SELINUS UNIVERSITY
OF SCIENCES AND LITERATURE

Development of a platform for Downstream Protein Pipeline process development with a focus on formulation aspects

By DMITRY SERBZHINSKIY

A DISSERTATION

Presented to the Department of
Biochemistry
program at Selinus University

Faculty of Life and Earth Science
in fulfillment of the requirements
for the degree of Doctor of Philosophy
in Biochemistry

2024

Declaration:

I do hereby declare that I am the sole author of this thesis and that its contents are only the result of the studies and research I have done. I also declare that I have not submitted this thesis to any other institution or university for any other degree.

Dmitry A. Serbzhinskiy, MS, MiT

Student ID: UNISE0632IT(UNISE0632UK). Selinus University

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to the team at Selinus University for their guidance and support throughout the entire process of preparing my thesis. My special appreciation goes to Dr. Salvatore Fava his invaluable support during my study period, always answering my questions and providing consultations without hesitation.

I extend my sincere thanks to the former and current SSGCID team, under the leadership of Dr. Peter J. Myler, for giving me the rare opportunity to work and learn alongside them. I am particularly grateful to Dr. Bart Staker, Dr. Isabel Phan, Sandhya Subramanian, Dr. Sally Lyons Abbott, Ariel Abramov, and Aarthi Sekar for their help, guidance, and support during my tenure at SSGCID.

Additionally, I appreciate the encouragement and support from my colleagues at KBio, including Barry Bratcher, Dr. Joshua Fuqua, Dr. Joshua Royal, Joshua Morton, and Steven Hume.

Finally, I would like to thank my family, Ira and Maria, for their unwavering patience and belief in me. Your support made this work possible. Thank you!

Table of Contents

Introduction

With the biotechnology field booming over the last several decades and significant breakthroughs occurring in various aspects of fundamental and applied science, it is difficult to determine which field should be the focus of one's research. However, based on my previous work, I believe that studying the applied aspects of protein engineering would be highly beneficial for the industry.

My research is focused on the platform approach for developing the downstream protein production pipeline. In this work, I summarize the current industry approaches, share my experience in the field, and outline future directions for the development of therapeutic protein products.

Classification of manufactured proteins based on current needs.

Proteins are complex organic macromolecules made from alpha-amino acids connected into the chains by peptide bonds. With the other biologic macromolecules such as nucleic acids they are true base of life on our planet. Known to humans through the entire history of our species, the scientific study of them began in the 18th century by Italian chemist Jacopo Beccari (1728) who first purified gluten from the wheat flour. Proteins were distinguished in a separate class of biomolecules later by Antoine de Fourcroy who noticed denaturation as an important feature of the proteins. Soon, in the early 19th century scientists discovered that the proteins were made from amino acids. The term "protein" was first proposed by Swedish chemist Jakob Berzelius. This word originated from Greek "πρωτος" – first, primary.

In the early 20th century, many fundamental discoveries have been made. German chemist Hermann Fischer proved that the protein is made from amino acid residues connected by peptide bonds, he was the first to made amino acid sequence analysis.

Great complexity of the functional protein molecules created (and continues to) great difficulty in their study and practical applications. First protein studies used abundant and easy to purify product such as blood proteins and digestive enzymes. Discovering secondary

and tertiary structures of proteins as well as developing sequencing and X-ray crystallography methods helped discover that the unique 3-D molecule structure is a key to understanding their function and opened a door to the modern field of protein engineering.

Although proteins were used as a food source, medicine, garment, and structural materials through the course of human history, later scientific advancements allowed the use of a bigger variety of them in much wider fields of human activity. We group them into the following categories:

Research and development product, Biotherapeutics, Food and Cosmetic Industry, New Composite Materials. Below we characterize each of those categories:

Research and development products

Currently, every new protein product manufactured undergoes the research and development phase to characterize the product and then choose a more efficient way to manufacture it. This is an important step before starting any biologics drug development process. For example, in the Structural genomic project where the author of this manuscript participated as a researcher, we need to express, purify, and characterize housekeeping proteins from different pathogenic microorganisms. For those purposes we did numerous experiments using tagged proteins, enzymes, antibodies and other types of commercially available products. Currently, companies like Thermo Fisher Scientific, and Abcam, play a crucial role in providing scientists necessary research tools. However, sometimes researchers need to produce custom proteins at their lab for specific research purposes or to save money.

Biological therapeutics

They are probably the most abandoned (by numbers and capital investments) class of manufactured proteins. It includes antibodies, enzymes, artificial hormones, antiviral and antibacterial proteins,

Protein therapeutics can be also grouped based on their molecular types including antibody-based drugs, anticoagulants, blood factors, bone morphogenetic proteins, engineered

protein scaffolds, enzymes, Fc fusion proteins, growth factors, hormones, interferons, interleukins, and thrombolytics (Carter, 2011)

Proteins used in the food and cosmetic industry.

Proteins are the central part of daily food products, however, obtaining balanced, healthy, and abundant protein sources has always been a challenge. With advances in agriculture including fertilization, mechanization, and genetic engineering humanity has better access to balanced food than ever before, however with the increasing world population and continued problem with malnutrition for most of the world, newer approaches to food manufacturing are very much actual. Multiple start-ups and research institutions are working on the development of a new generation of food products and additives. One such example is using RuBisCo (Ribulose bisphosphate carboxylase/oxygenase enzyme) as a food additive. This photosynthetic complex is probably the most abandoned protein in plants, at the same time it is a protein containing all essential amino acids which makes it a perfect candidate for the biomanufacturing of food additives. Other examples can be legume proteins which are also abundant in plants and can be easily purified on commercial scales.

New composite materials

There are several materials used as adhesive and sealants, for surface coating and gap filling, as well as several fibrous materials related protein products. However, because they are not usually expressed and purified using modern biotechnology methods, they are not subject to this work.

Literature review

Stages of modern protein production

Protein production is a complex multistage process including the engineering of a sequence with desired properties, multiple molecular biology steps to introduce the sequence into expressing host, expression, purification, formulation, and future potential upscaling of the process if this is a commercial entity. Below we will describe the process in greater detail.

Research and development

All scientific experiments start with a problem statement, in the case of protein production we need to answer questions: what product do we need to make, what its properties, how pure the product will be, how much product do we need, what methods do we use, what would be an endgame for the product cycle?

Genetic engineering and Expression

It is possible to produce simple proteins by chemical ligation, but a much more efficient method is to make cells of living organisms to produce the necessary molecule.

The genetic manipulation of organisms is the modification of genetic material through various techniques with the goal of making the organism produce a non-native molecule. The technique was first accomplished in 1973 when a gene for frog ribosomal RNA was transferred into *Escherichia coli (E.coli)* cells and successfully expressed the foreign gene in multiple bacterial colonies (Cohen et al., 1973). This was followed by the integration of retrovirus DNA (Moloney leukemia virus) into mouse embryos that developed leukemia to prove that inserted transgenes can be passed onto future generations (Jaenisch, 1976). Then, commercial protein production using transgenic organisms began to rise with the development of the hormone somatostatin in 1977 (Houghes, 2011), human insulin in 1978 (Crea et al., 1978), and human growth hormone in 1979 (Ayyar, 2011). All those proteins were expressed in E. coli by the biotechnology company Genentech. With these technological advancements, select yeast, bacteria, and mammalian organisms were commonly chosen to continue the development of genetic engineering. These traditional expression systems have

been used to produce many recombinant protein products which have contributed to the large market size of $1.6 billion in 2017 (MarketandMarkets(™), 2017). Recently, the production of recombinant proteins from traditional expression systems has expanded focus to plant systems with the first Food and Drug Administration (FDA) approved recombinant protein from a transgenic plant in 2012. This product was known under trademark ELELYSO(™)

Expression hosts

A wide variety of expression systems have been used to produce engineered proteins, currently however, about 3/4 of approved recombinant proteins are produced in Chinese hamster ovary (CHO) cells (Butler and Spearman, 2014). However, using those cells for protein expression has certain downsizes. Below we will review the main expression systems currently used for protein expression as well as the advantages and disadvantages of each system.

Bacteria

Escherichia coli (E. coli) is a preferred organism for producing recombinant proteins. Its use as a cell factory is well-established, making it a popular expression platform. Consequently, many molecular tools and protocols are available for high-level protein production. These tools include a variety of expression plasmids, numerous engineered strains, and diverse cultivation strategies.

The advantages of using E. coli as the host organism are well known. It has unparalleled fast-growth kinetics; in glucose-salt media and under optimal conditions, its doubling time is about 20 minutes (Sezonov et al., 2007). This means a culture inoculated with a 1/100 dilution of a saturated starter culture can reach the stationary phase in a few hours. However, expressing a recombinant protein can impose a metabolic burden on the bacteria, leading to a decrease in generation time (Bentley et al., 1990). Despite this, high cell density cultures are easily achieved. The theoretical density limit of an E. coli liquid culture is

estimated to be about 200 g of dry cell weight per liter, or roughly 1 × 10^13 viable bacteria per milliliter (Lee, 1996; Shiloach and Fass, 2005).

In complex media, bacterial growth density is much lower than this theoretical limit. In the simplest laboratory setup, the upper limit is less than 1 × 10^10 cells per milliliter (Sezonov et al., 2007), which is less than 0.1% of the theoretical limit. To address this, high cell-density culture methods were developed to boost E. coli growth, even during recombinant protein production (Choi et al., 2004). Additionally, transforming E. coli with exogenous DNA is fast and easy, allowing plasmid transformation to be performed in a short time.

The main challenges in designing E. coli-based protein production pipelines include finding or designing the correct plasmid, selecting an appropriate promoter, choosing suitable affinity tags and tag removal features, and identifying the correct bacterial strains. When expressing various eukaryotic proteins, researchers may encounter several challenges that could strain a research or manufacturing project. These issues include no or low expression, protein toxicity, codon bias, and misfolding and aggregation of expressed proteins. Below, we will address each of these problems and suggest optimal solutions.

Yeast as expression system

Yeasts, such as Saccharomyces cerevisiae (S. cerevisiae) and Pichia pastoris (P. pastoris), are ranked by many industry experts as the second most employed systems for generating recombinant proteins (Butler and Spearman, 2014). This popularity stems from their ability to achieve high yields through fermentation, along with their capacity for post-translational modifications that resemble eukaryotic glycosylation.

As protein production hosts, yeasts offer significant advantages, including proper post-translational modifications, rapid growth, straightforward genetic manipulation, scalable fermentation, high biomass concentrations, and pathogen-free production. Yeast expression systems can be categorized into two groups based on their metabolism: methylotrophic and non-methylotrophic. Due to these features, yeast expression systems are among the most

frequently utilized eukaryotic organisms, serving as models for studying gene expression regulation, signal transduction, aging, apoptosis, metabolism, cell cycle control, the secretory pathway, and numerous other crucial biological processes.

Recombinant proteins can be expressed intracellularly or directed to the secretory pathway using a signal peptide. A commonly employed signal sequence functional in all yeast expression systems is the prepro-sequence of mating factor α1 (MFα1). Additionally, *Saccharomyces cerevisiae* holds key safety advantages, being generally regarded as safe (GRAS) due to its nonpathogenic nature and historical use in various nutritional industries and the production of biopharmaceuticals. The current understanding of yeast genetics, physiology, and fermentation further supports its application in producing valuable products.

Notably, products derived from S. cerevisiae, such as hirudin, insulin, glucagon, macrophage colony-stimulating factor, and platelet-derived growth factor, are currently available on the market. Despite these successes, the commercial utilization of S. cerevisiae has been constrained by factors such as hyper glycosylation of proteins, low protein yield, and plasmid instability. These limitations have prompted the exploration of alternative expression systems, including methylotrophic yeasts like *Pichia pastoris* and *Hansenula polymorpha*, as well as non-methylotrophic yeasts such as *Yarrowia lipolytica, Kluyveromyces lactis*, and *Arxula adeninivorans*.

Pichia pastoris stands out as an excellent host for producing heterologous proteins, including industrial enzymes and biopharmaceuticals. This methylotrophic expression system has proven successful in generating various recombinant proteins, including human erythropoietin, phospholipase C, phytase, human superoxide dismutase, trypsin, human serum albumin, collagen, and the human monoclonal antibody 3H6 Fab fragment. The expanded use of *P. pastoris* has led to the development of novel genetic tools, increasing the availability of new strains and, consequently, heightening the demand for improved cultivation and production procedures with this yeast. Notably, *P. pastoris* demonstrates superior efficiency in the secretory production of recombinant proteins compared to other yeast species (Zha et al., 2023).

Enhancing the efficiency of heterologous protein production through host strain engineering in yeast cells is a costly and time-consuming process. An alternative, more economical approach involves optimizing culture conditions using experimental design strategies (Asada et al., 2011). Various factors such as pH, oxygen density, temperature, aeration, and induction techniques can significantly impact yeast culture yields. Optimal conditions for protein expression in *P. pastoris* vary depending on the host strain and the proteins being expressed. These conditions can include different medium compositions, varying pH levels, temperatures, oxygen densities, and methanol concentrations.

Plant expression systems

Plant expression systems encompass various platforms with distinct advantages and drawbacks. Bacteria, notably E. coli, represent a well-established and extensively researched expression platform with molecular tools facilitating expression levels of up to 30% of total cellular protein (Rosano and Ceccarelli, 2014; Baeshen et al., 2015). However, bacteria lack the capability for protein glycosylation, which is crucial for the biological activity of many proteins. Consequently, expressed proteins often accumulate in inclusion bodies, necessitating additional steps such as solubilization and refolding, leading to significant reductions in protein yields (Singh et al., 2015). Moreover, many complex and heterologous proteins simply cannot be expressed in bacteria due to the lack of proper cellular machinery.

Reducing downstream processing expenses can be achieved by secreting recombinant proteins into the medium. This strategy simplifies purification and reduces the need for extensive cell lysis and protein extraction processes. Additionally, additives in upstream processing can boost yields without significantly increasing production costs, primarily through improvements in cell line selection, media optimization, and expression levels (Gronemeyer et al., 2014). Techniques such as metabolic engineering and the optimization of fermentation conditions can also contribute to higher yields and more cost-effective production processes.

Notably, the cost of producing recombinant proteins using bacteria is considerably lower than with CHO cells and is comparable to plant platforms. Bacterial systems benefit from

relatively inexpensive growth media and rapid growth rates, which reduce the overall time and cost of production. However, the limitations in protein folding and post-translational modifications often necessitate the use of eukaryotic systems, such as yeast and plant expression platforms.

In recent years, protein expression levels in plant systems have surged, now rivaling or even surpassing those in bacterial and yeast systems. Plant systems exhibit the capacity to generate complex proteins while maintaining their bioactivity through post-translational modifications (Streatfield, 2007). These modifications include glycosylation patterns that closely resemble those in humans, making plant-produced proteins suitable for therapeutic applications.

Plant systems are categorized into leafy, seed, and bioreactor systems, each presenting unique advantages and disadvantages. Leafy systems, such as those using tobacco plants, offer rapid biomass accumulation and are ideal for transient expression of proteins. Seed-based systems provide long-term storage of recombinant proteins within the seed matrix, which can be advantageous for stability and ease of harvest. Bioreactor systems, which utilize plant cell cultures in controlled environments, offer high levels of control overgrowth conditions and protein production.

In general, plant systems efficiently express recombinant proteins and offer scalability from pilot to large-scale production processes (He et al., 2011). The ability to scale up production without significant increases in cost is a critical advantage, particularly for the production of pharmaceuticals and industrial enzymes. As research continues to improve plant expression systems, their role in biotechnology is expected to expand, offering a versatile and cost-effective alternative to traditional microbial and mammalian cell systems.

Algae expression system

Algae have been advocated as an alternative platform to higher plants due to their solar-powered nature and proficiency in properly folding and assembling complex animal proteins. The capacity for sexual reproduction in higher plants introduces breeding

strategies not feasible in other ex vivo systems, enabling the selection of valuable traits. However, the potential for gene flow into non-genetically modified plants poses a regulatory challenge for developing recombinant therapeutics. Moreover, higher plants allocate resources to tissue and organ development, leading to inefficient space utilization. Despite these drawbacks, some edible plant species offer opportunities for the administration and delivery of therapeutic proteins, substantially reducing post-production processing and purification costs.

Eukaryotic single-celled microalgae address challenges posed by the mentioned systems while retaining numerous advantages. Their lack of cell differentiation or tissue development results in space-efficient uniformity utilized in cell culture systems. Additionally, microalgae can reproduce sexually, providing breeding opportunities like higher plants but within controlled environments like photobioreactors. Microalgae exhibit unprecedented genetic diversity compared to domesticated terrestrial crops, offering untapped environmental tolerance and disease resistance traits beneficial for large-scale production. Advances in genetic engineering and synthetic biology enable the transfer of these traits between microalgae species.

Microalgae present a potential for scale and cost of production that could rival traditional agricultural methods once the platform is refined for efficient recombinant protein production. The current therapeutic protein market is dominated by mammalian cell culture, but the resulting products are often only accessible to those who can afford their exorbitant prices. To illustrate, the estimated cost of monoclonal antibody production is around $150 per gram in mammalian cells, whereas it is only $0.05 per gram in plants (Dove, 2002). Additionally, establishing mammalian cell culture production facilities incurs substantial upfront costs, reaching several hundred million dollars (Dove, 2002). Microalgae emerge as a promising alternative due to their cost-effective cultivation, with media costs as low as $0.002 per liter, and production facilities that can be a fraction of the expense of a mammalian cell culture facility. This holds particular significance for recombinant proteins required in large, affordable quantities, such as animal feed, industrial enzymes, or vaccines for developing countries.

Higher plants as expression system

Higher plant expression systems, as mentioned earlier, offer several potential advantages over conventional methods and have demonstrated their reliability in producing highly valuable proteins (Lindsay et al. 2018; Malm et al. 2019; Ward et al. 2021). The first instance of biopharming in plants dates to 1986 when a chimeric human growth hormone was produced in transgenic tobacco and sunflower callus tissue (Barta et al. 1986). Notably, the FDA-approved taliglucerase alfa, a genetically modified plant-derived therapeutic, was produced in transgenic carrot cell suspension cultures for treating Gaucher disease (Zimran et al., 2011). Utilizing bioreactors, this production system claims lower initial investment and running costs compared to mammalian-based systems (Tekoah et al., 2015)

Among various plant expression systems, *Nicotiana benthamiana* serves as the core production host for numerous companies, including KBio, Icon Genetics, iBio, and UniBio. This plant, exhibits remarkable susceptibility to pathogens, making it an excellent host for transient gene expression (Bally et al. 2015). In experiments, *N. benthamiana* plants are typically grown for 4–7 weeks and then infected with Agrobacterium tumefaciens carrying the genes of interest (GOI). The peak level of GOI product is usually reached 3–7 days after infection.

Recent studies demonstrated that crude extracts of *N. benthamiana* leaves transiently expressing influenza virus hemagglutinin (H5) trimers prevented the spread of avian influenza (H5N1) when injected into chickens (Phan et al. 2020). Additionally, randomized phase 3 trials showed efficacy, safety, and immunogenicity in humans for a quadrivalent virus-like particle influenza vaccine produced in *N. benthamiana* by Medicago Inc. (Quebec, Canada) (Tregoning 2020). This marks a crucial advancement in plant-derived biologics, promising more accurate protection and a broader range of production options as next-generation vaccines enter the market.

It is evident that higher plants offer numerous benefits for recombinant protein production. Plants, being distinct from animals, pose a low risk of contamination and replication of

human pathogens within the system (Commandeur and Twyman 2004). The simplicity of cultivation, which doesn't require a sterile environment, coupled with inexpensive fertilizer solutions (<0.002 $/L), contrasts with the costly media needed for mammalian cell culture, exceeding 50 $/L (Xu et al. 2017). Moreover, higher plant expression systems reduce costs for purifying the protein of interest and testing for virus-free status. Using *A. tumefaciens* and/or viral vector-mediated transient expression system, recombinant protein expression in plants can be achieved approximately 8 weeks after obtaining the corresponding DNA sequence (Gleba et al. 2014; Shoji et al. 2012). Additionally, plant expression systems excel in producing intrinsically disordered proteins (Gengenbach et al. 2019), which may not be efficiently synthesized in mammalian cells or prokaryotes due to toxicity or complex structure. Furthermore, plant expression systems have the advantage of producing much larger recombinant proteins compared to bacterial expression systems.

In certain instances, the biological activity of recombinant proteins relies on proper protein folding. Prokaryotic expression systems may face challenges in achieving correct protein folding due to limitations in bacterial protein processing complexes and posttranslational modification capacities (Sahdev et al. 2008). Conversely, plants possess the capability to assemble and perform posttranslational modifications on large multimeric proteins. However, it's important to note that glycosylation mechanisms vary among species, posing a significant concern for any non-human expression system. Given that humans regularly encounter plant glycoproteins in their diet, glycosylated proteins produced in plants may be deemed acceptable for topical and oral administration (Gomord et al. 2005).

Apart from *N. benthamiana, N. tabacum* is employed for both stable and transient expression of recombinant proteins. Additionally, various crops, fruits, and vegetables, including rice, maize, lettuce, tomato, and potato, have undergone evaluation to produce recombinant proteins (Shanmugarai et al. 2020).

There are two categories of transient protein expression systems in higher plants. One relies on plant viruses, such as the tobacco mosaic virus (TMV), while the other utilizes the Agrobacterium-mediated transient gene expression system, commonly known as agroinfiltration (Burnett and Burnett 2020). The virus-based system carries the potential

risk of viral vectors infecting plants in the ecosystem, as plant viruses replicate autonomously. In the agroinfiltration method, a suspension of *A. tumefaciens* is introduced into plant leaves through injection or vacuum infiltration. In this process, the bacteria facilitate the transfer of the gene of interest (GOI) into the nucleus of host plant cells by delivering T-DNA. Generally, the expression of recombinant proteins via agroinfiltration surpasses and is more efficient than that achieved through traditional plant transformations.

Insect cells as expression system

The utilization of Baculovirus-mediated expression in insect cells has become firmly established to produce recombinant glycoproteins. This method is favored due to its ease and speed in expressing heterologous proteins on a laboratory scale, along with a high probability of obtaining biologically active proteins. While *Spodoptera frugiperda* Sf9 cells are widely used in this approach, other primarily lepidopteran cell lines are also employed for protein expression (Altmann et al., 1999). Although recombinant baculovirus is the conventional vector for foreign gene expression, stable transfection of insect cells, especially dipteran cells, presents an intriguing alternative. Insect cells can be cultivated on serum-free media, offering cost and biosafety advantages. Conditions for large-scale culture have been developed to meet the specific requirements of insect cells. In terms of protein folding and post-translational processing, insect cells are second only to mammalian cell lines, as evidenced by numerous processing events occurring in both systems.

Insect cells require the use of an intermediate, specifically the baculovirus, for protein expression. Baculoviruses, a diverse group of DNA viruses, can infect various insect cells, and act as shuttles for introducing the target gene into a given host cell. *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) is the most characterized and widely used baculovirus for this purpose (Ros., 2020). The process involves inserting the gene encoding the protein of interest into a primary vector, which is then cloned into a secondary vector called "Bacmid." The Bacmid is transferred into a bacterial strain (usually E. coli) for initial virus production and assembly, resulting in the generation of the first baculovirus (P1). The

P1 virus is amplified in an insect cell (e.g., sf9) to achieve a suitable titer (P2), and the P2 virus is then used to infect the same or a different insect cell line for protein expression.

Insect cells serve as versatile expression hosts for a variety of recombinant proteins, thanks to their strong folding capability and relatively high culture density. This makes them excellent choices for expressing complex intracellular and virus proteins. In 2007, Cervarix®, an HPV vaccine produced by an insect cell line in the format of virus-like particles (VLPs), was approved for human use (Zimmerman., 2007). Beyond their potential in therapeutic and vaccine development, highly active proteins produced in insect cells find widespread applications in various disciplines, such as biophysics and biochemistry, for structure elucidation, drug design, assay establishment, and diagnostic reagent development.

Mammalian expression systems

Over the past decade, protein therapeutics derived from mammalian cells have significantly changed the landscape of human healthcare. The rising importance of protein therapeutics has driven efforts to develop more cost-effective and efficient cell lines capable of producing high-quality protein products. Mammalian cell-based bioprocesses have been extensively employed in the production of viral vaccines, diagnostic tools, and therapeutic proteins. These cells serve as hosts for protein production in the manufacturing of protein therapeutics (Wurm., 2004). Among the most widely used host mammalian cells are Chinese hamster ovary (CHO) cells and mouse myeloma cells. The two most common cell lines in bioprocessing today, DUKX-X11 and DG44 (Khan., 2013), originated from derivatives of the CHO cell line—CHO-K1 and CHO pro-3—which were engineered to be deficient in dihydrofolate reductase (DHFR) activity.

Various mammalian cell lines have been employed for protein expression, with HEK 293 (Human embryonic kidney) and CHO (Chinese hamster ovary) being the most prevalent. These cell lines can undergo transfection using methods such as polyethyleneimine (PEI) or calcium phosphate. HEK 293 cells exhibit high levels of PEI-mediated transfection, with 50–

80% of cells expressing green fluorescent protein (GFP). Consequently, HEK 293 cells are widely utilized for producing recombinant proteins through both transient transfection and the establishment of stable cell lines (Khan., 2013).

Protein expression in mammalian cells can also be achieved through viral-mediated transduction. This approach utilizes recombinant baculoviruses for straightforward transduction of mammalian cells, enabling the production of milligram quantities of proteins for structural studies. Other cell lines, including COS and Vero (both from green African monkey kidney), HeLa (Human cervical cancer), and NS0 (Mouse myeloma), have also been employed in structural studies. Certain cell lines, such as NS0, are more challenging to transfect and typically require methods like electroporation, which are primarily used in the production of stable cell lines.

The main advantages of mammalian cell expression lie in their ability to properly and efficiently recognize signals for the synthesis, processing, and secretion of eukaryotic proteins. This capability ensures that the proteins produced are correctly folded and post-translationally modified, which is crucial for their biological activity and therapeutic efficacy. However, it is essential to note that there are species differences that should be considered, as these can impact the efficiency and quality of protein production.

Overall, the continued advancement and optimization of mammalian cell lines and bioprocesses are vital for meeting the growing demand for protein therapeutics. By enhancing the efficiency and cost-effectiveness of these systems, the biotechnology industry can better address the needs of patients and healthcare providers, ultimately improving outcomes and access to life-saving treatments.

Utilization of Mammalian Expression System:

The initial biologic approved from a mammalian bioprocess platform was tissue plasminogen activator (tPA), introduced by Genentech Inc. in 1987 (Wuest et al., 2012). Currently, the production of biologics in mammalian cells holds a predominant position. Out of the 58 biopharmaceutical products approved between 2006 and 2010, 32 were

manufactured in mammalian cells, 17 in E. coli, four in yeast, three in transgenic animals, and two in insect cultures.

Limitations of Mammalian Expression System:

While the mammalian expression system is recognized for providing functional proteins due to glycosylation, it comes with certain limitations. Despite the biological activity retained in proteins obtained through gene expression in mammalian cells, the system is acknowledged for its high cost. The intricate technology and potential contamination with animal viruses in mammalian cell expression pose challenges for large-scale industrial production. Nevertheless, the system is extremely valuable for expressing various heterologous proteins, including viral structural proteins and bioactive peptides, for specific functional analysis.

Cell free expression systems

Another method gaining popularity among researchers involves the utilization of isolated cell lysates. These lysates contain all the essential machinery for protein synthesis while lacking cell debris, genomic DNA, and proteases. Known as cell-free expression systems, this approach has garnered attention in the scientific community. One of the illustrations of this technology presented in Fig. 1

Over the past few years, there has been a notable rise in the accessibility of novel cell-free lysates sourced from various organisms, facilitating their application in synthesizing a broad spectrum of proteins. However, significant hurdles persist concerning scalability, cost efficiency, protein folding, and functionality despite these advancements.

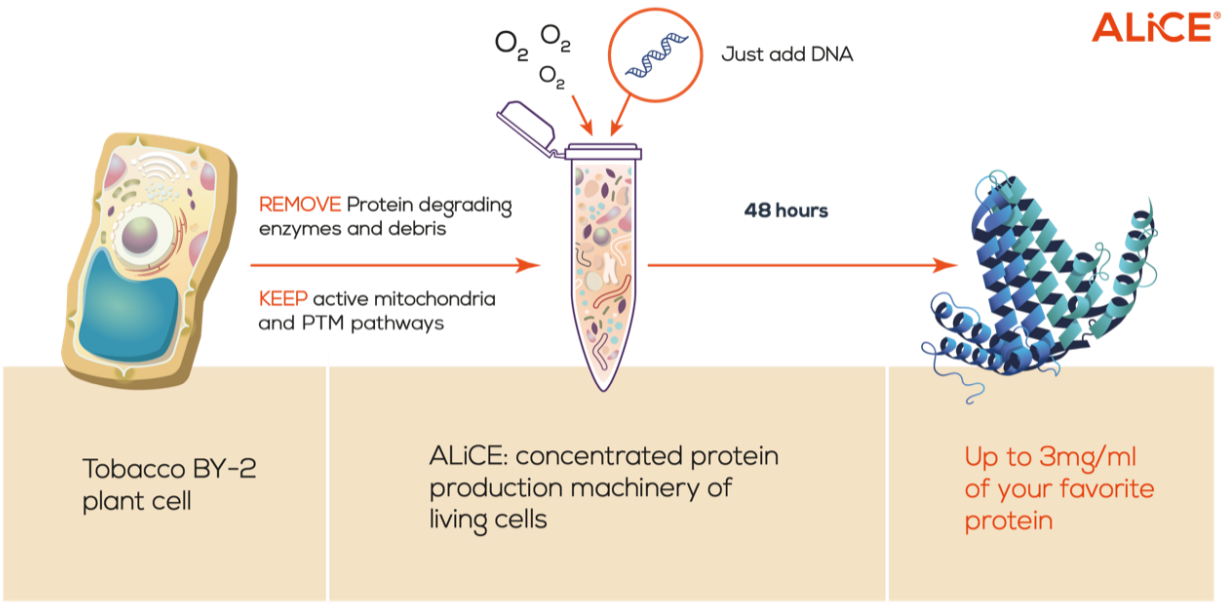


Figure 1. AliCE Cell-free expression system based on Nicotiana tabacum cell lysate. Source: LenioBio.com.

Protein Purification

Protein purification is generally a set of steps aimed to isolate one or a few proteins from different mixtures, usually taken from cells, tissues, or even whole organisms. The process is obviously important to research protein functions and structure. Purification sorts out the protein molecules from the contaminants in the mix and finally, taking the protein of interest, leaving all the others behind. This part, maybe the most difficult and long during entire protein production process. Different protein properties can be used to implement this process. It can be molecule size, charge, hydrophobicity etc.

When it comes to protein purification, there are two main avenues: preparative and analytical. Preparative purifications are all about getting amount of pure protein ready for later use. This could be for making enzymes, nutritional proteins, or certain therapeutics. Usually, purification scientists must run through many of steps to make sure the final protein product is free of contaminants. On the other hand, analytical purifications could give a

scientist just a small amount of protein for further testing. Those tests can include identification, quantification, and studies of the protein's structure, post-translational modifications and function. Pepsin was the first protein purified to the point that they could be crystallized (Rawlings and Salvesen, 2013).

Extraction

If the protein of interest is not naturally released by the expressor cell into its surrounding, the first purification step involves breaking the cells that contain it. Depending on how delicate the protein is and how tough the cells are, the following methods can be used: 1. freezing and thawing repeatedly, 2. using sound waves (sonication), 3. applying high pressure to break the cells, 4. grinding the cells (for example with bead mill), or 5. making the cell membranes more permeable using detergents and/or enzymes. Afterward, the solid part of cell debris can be separated by spinning the mixture in a centrifuge, leaving the proteins and other soluble components in the liquid supernatant.

When cells are destroyed, proteases are released into the mixture, and can start digesting proteins of interest. If the protein is sensitive to proteolysis, it's important to work quickly and keep the extract cold to slow down this process. Alternatively, one or more protease inhibitors can be added to the mixture before cell lysis. Sometimes, in cases of the samples containing significant amount of DNA nucleases should be added to the mixture to decrease viscosity of it.

Filtration

After cell/tissue homogenization, the next step in protein purification often involves removing cellular debris, organelles, and insoluble components to isolate the soluble fraction containing the proteins of interest. This initial separation step is typically achieved through filtration and/or centrifugation. Filtration is a ubiquitous separation technique utilized across various industries, for purifying substances based on their physical properties, usually the size of the mixture's components. At its core, filtration involves passing a fluid

mixture through a porous barrier, the filter medium, which selectively retains particles larger than its pore size while allowing smaller particles and solvent to pass through. This process effectively separates components based on disparities in size, shape, and charge. In the realm of protein purification, filtration methods are indispensable for isolating and concentrating target proteins from complex biological matrices (see Fig. 2.)

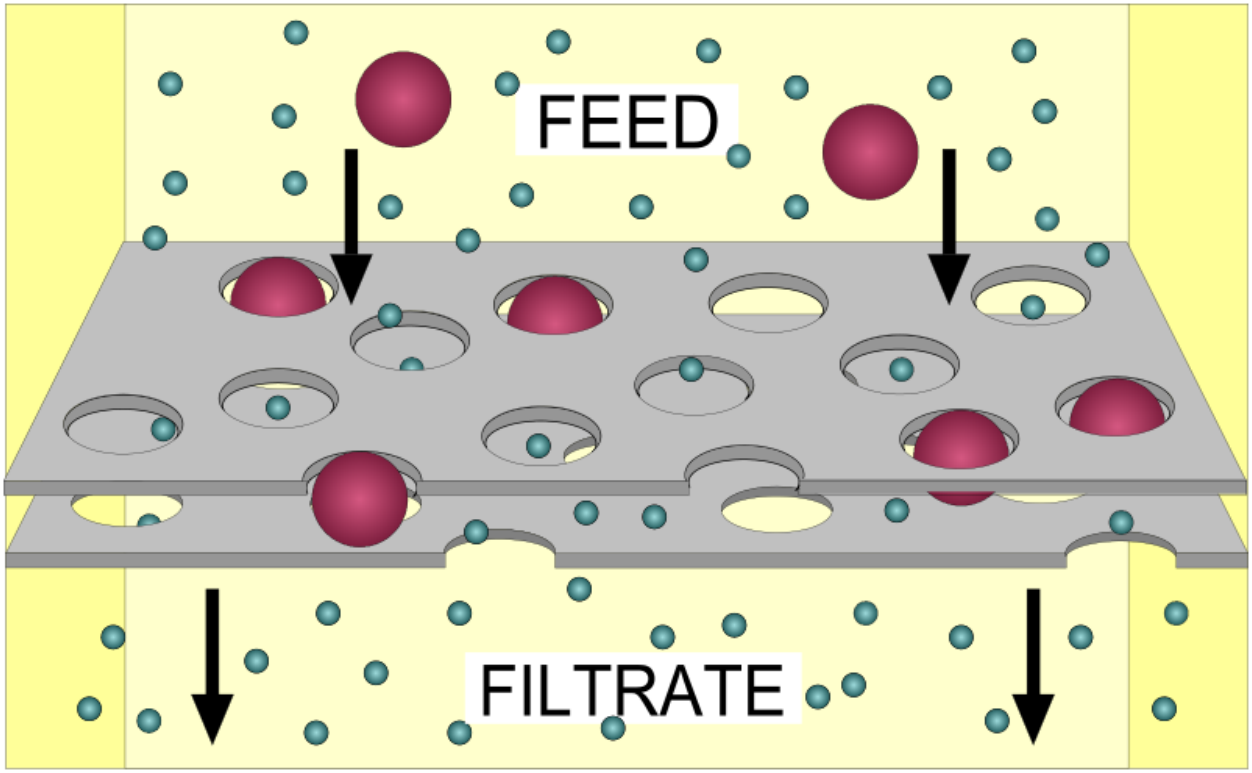


Figure 2. Schematics of filtration process (Source: Wikipedia)

While speaking about filtration related to protein purification, we cannot avoid talking about Tangential Flow Filtration (TFF), also known as crossflow filtration, is a versatile and widely used technique in bioprocessing and protein purification. Unlike traditional filtration methods where the fluid flows directly through the filter medium, in TFF, the fluid flows tangentially across the surface of the filter membrane. This creates a shearing force that constantly sweeps away retained particles, allowing for continuous filtration without clogging (Fig. 3).

A common TFF system consists of three main components: a filter membrane and a holder, a feed stream containing the solution to be filtered, and a permeate stream where the filtrate

flows through. The feed stream is pumped parallel to the membrane surface, and as the fluid flows across the membrane, the smaller molecules pass through the pores while larger molecules, such as proteins or particles, are retained (Fig. 3). This differential separation enables the concentration and purification of proteins from complex mixtures based on size, allowing for the removal of impurities and smaller molecules.

One of the significant advantages of TFF is its scalability and flexibility, making it suitable for various applications from laboratory-scale research to large-scale industrial processes (Fig. 4). It offers precise control over process parameters such as flow rate, pressure, and membrane pore size, enabling optimization for specific purification needs. TFF is commonly used for concentration, buffer exchange, desalting, and diafiltration of proteins, providing high yields and purity while minimizing sample handling and loss. Overall, TFF has revolutionized protein purification by offering efficient, continuous, and scalable separation methods essential for biopharmaceutical production and biomedical research.

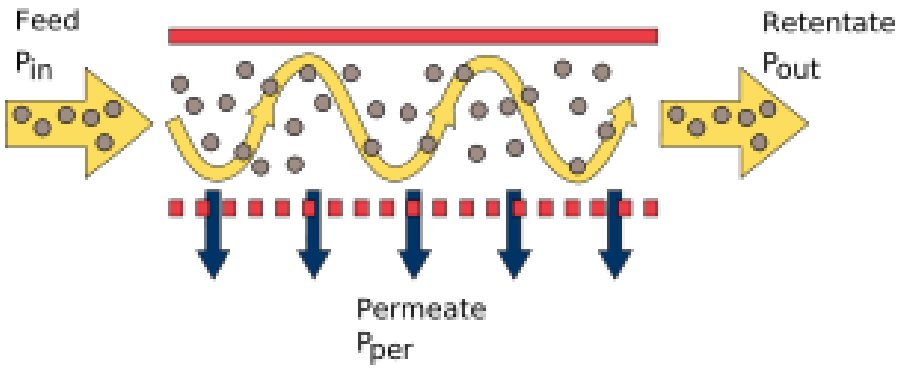


Figure 3. Diagram of Tangential Flow Filtration (Wikipedia)

Figure 4. Ceramic membrane for Industrial TFF (Wikipedia)

Solubilization/Precipitation strategy

In large-scale protein purification, a common initial step to separate proteins involves precipitation using a salt like ammonium sulfate ($(NH_4)_2SO_4$). This technique is known as Salting In or Salting Out (refer to Fig. 5). It works by gradually adding increasing amounts of ammonium sulfate to the solution and then collecting the different fractions of protein that precipitate out. Ammonium sulfate is a preferred choice because it dissolves well in water, is relatively insensitive to temperature changes, and is generally non-damaging to most proteins. Additionally, it can be easily removed through dialysis (Fig. 6). During this process, the hydrophobic regions of the proteins become exposed, attracting other hydrophobic regions of other proteins and thus, causing aggregation. The precipitated protein particles become large enough to disperse the light and be visible. One of the advantages of this method is its cost-effectiveness, especially for the large volumes.

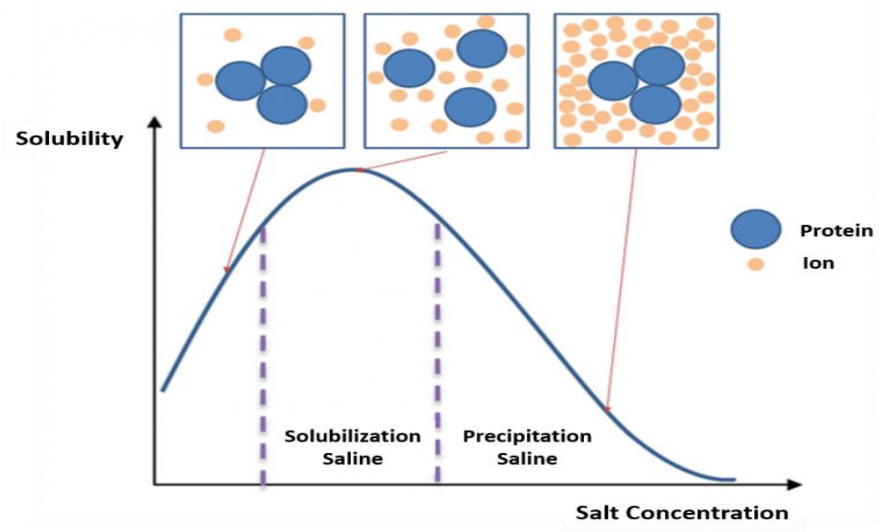


Figure 5. Salting In and Salting Out. During the salting in process, salt molecules increase the solubility of proteins by reducing the electrostatic interactions between protein molecules.

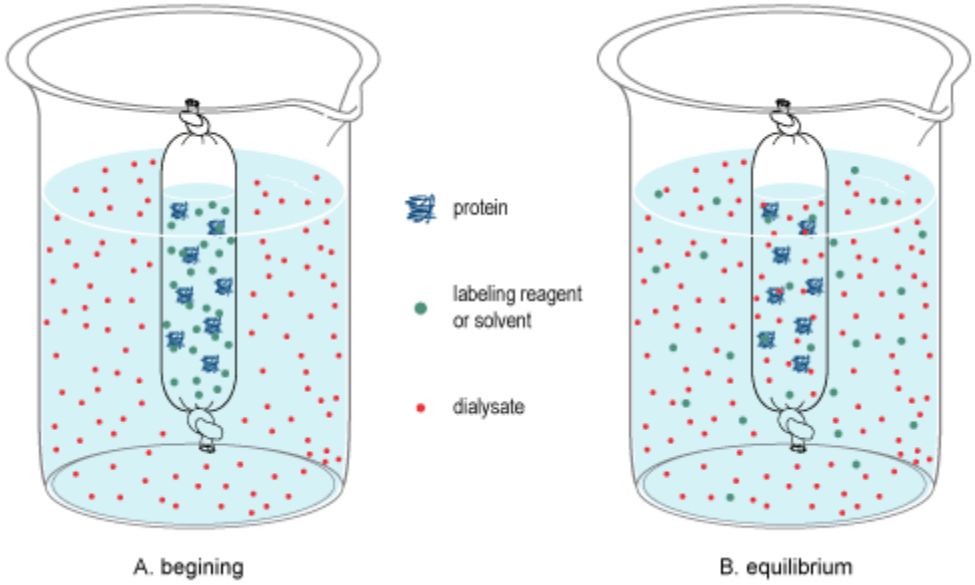


Figure 6. Dialysis.

The process of dialysis separates dissolved molecules by their size. The biological sample is placed inside a closed membrane, where the protein of interest is too large to pass through the pores of the membrane, but through which smaller ions can easily pass. As the solution comes to equilibrium, the ions become evenly distributed throughout the entire solution, while the

protein remains concentrated in the membrane. This reduces the overall salt concentration of the suspension. (Source for fig. 5 and 6: Wikipedia)

Centrifugation

Centrifugation is a technique that employs centrifugal force to segregate mixtures of particles with different masses or densities suspended in a liquid. When a vessel, often a tube or bottle, containing a mixture of proteins or other particles such as bacterial cells, is rotated rapidly, the inertia of each particle generates a force aligned with its velocity, which is proportional to its mass. This force is counteracted by the resistance exerted by the liquid on the particle, affecting its movement through the liquid. As a result of centrifugation, particles of varying mass or density move outward at different rates, with heavier, smaller, and denser particles moving faster than less massive particles or those experiencing more resistance in the liquid. This separation often leads to the formation of a concentrated "pellet" at the bottom of the vessel, enriched with the most massive particles having low drag in the liquid.

Particles that do not compact together remain predominantly in the liquid phase, referred to as the "supernatant," and can be separated from the vessel, effectively separating the supernatant from the pellet. The speed of centrifugation is determined by the angular acceleration applied to the sample, typically measured in terms of g-force. With sufficient duration of centrifugation, particles in the vessel reach equilibrium, where they accumulate at a specific point in the vessel where their buoyant density is balanced with the centrifugal force. This equilibrium centrifugation process facilitates extensive purification of specific particles.

General protein purification strategy

At the beginning of any purification, we need to note, that selection of an initial material is critical in devising a purification strategy. Typically, in plants or animals, a specific protein

isn't evenly spread throughout the body; instead, various organs or tissues exhibit different concentrations of the protein. Utilizing only the tissues or organs with the highest concentration reduces the volumes necessary to generate a certain quantity of purified protein. If the protein is scarce or holds significant value, researchers might employ recombinant DNA technology to engineer cells capable of producing large quantities of the desired protein (known as an expression system and explained in the above sections). Through recombinant expression, the protein can be tagged, i.e., certain affinity fragments could be attached to the protein molecules. Among widely used tags are 6-Histidine-tag and Strep-tag (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys), they are facilitating the purification process and also can reduce the number of purification steps required (Schmidt and Skerra., 2007).

Analytical purification typically exploits three characteristics to segregate proteins. First, proteins may be purified based on their isoelectric points by subjecting them to a pH gradient gel or an ion exchange column. Second, proteins can be separated according to their size or molecular weight using techniques like size exclusion chromatography or SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis). Third, proteins may be separated based on their hydrophobicity through methods like reversed-phase chromatography.

In preparative protein purification, the typical purification process involves one or multiple chromatographic steps. The fundamental concept of chromatography entails passing the protein-containing solution through a column filled with different materials. Due to varied interactions with the column material, different proteins can be separated based on the time they take to pass through the column or the conditions necessary to release the protein from the column. Typically, proteins are identified as they exit the column and passing through number of sensors which collected different types of data, most importantly UV absorbance at wavelength 280 nm. Numerous chromatographic techniques are available, with the most prevalent ones will be described below.

Size Exclusion (AKA Gel Filtration) Chromatography (SEC).

SEC It is an isolation method that involves the use of beads that have "tunnels" in them, and each have a precise size (Fig. 7). The size is referred to as an "exclusion limit," which means that molecules above a certain molecular weight will not fit into the tunnels. Molecules with sizes larger than the exclusion limit do not enter the tunnels and pass through the column relatively quickly by making their way between the beads. Smaller molecules, which can enter the tunnels, do so, and thus, have a longer path that they take in passing through the column. Because of this, molecules larger than the exclusion limit will leave the column earlier, while smaller molecules that pass through the beads will elute from the column later. This method allows separation of molecules by their size.

SEC is an ideal method for samples that are sensitive to changes of buffer conditions such as metal ion concentration, pH value, *etc*. To perform SEC, a packed bed is established by packing the size exclusion medium into an SEC column. The packed SEC column is equilibrated with SEC buffer to maintain the space between the medium particles and the pores of the matrix were totally filled with SEC buffer. Samples are injected into the column and then they are flowing pass the packed bed, during which the separation occurs. Components of samples are eluted isocratically, i.e. the composition of the mobile phase (the solvent moving through the column) remains constant throughout the entire run. SEC can be performed in the cold room or 37°C to meet the requirements of experiments. Biological macromolecules can be collected in any SEC buffer with the presents of cofactors, detergents, essential metal ions, denaturants, *etc*.

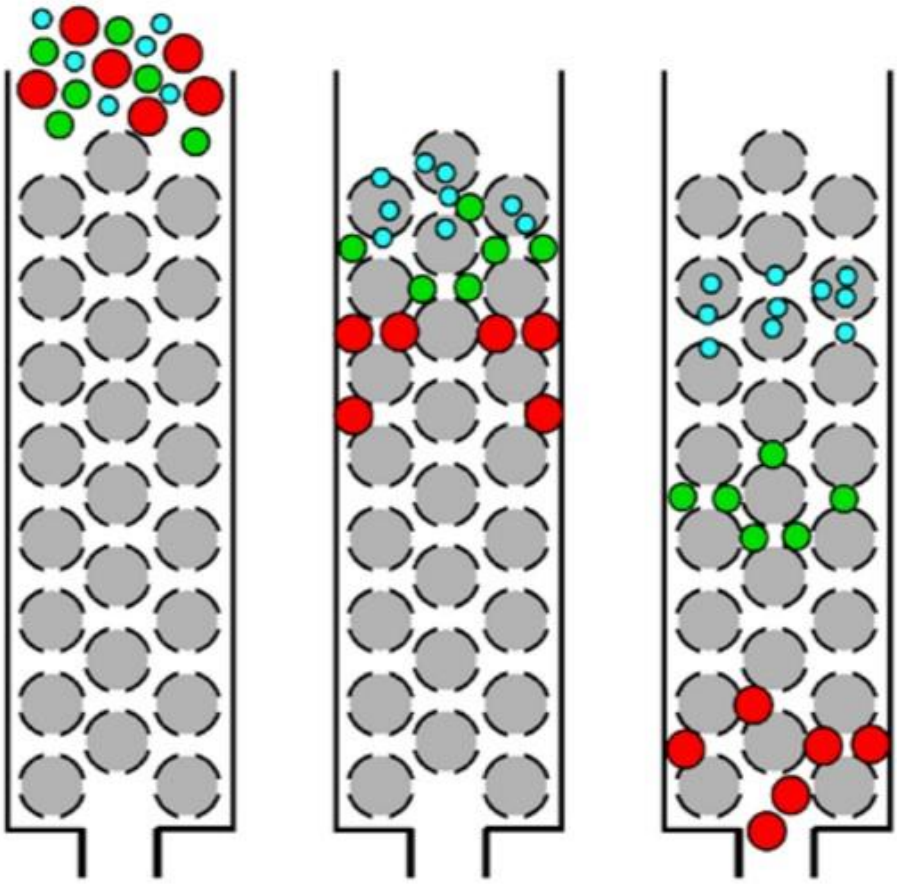


Figure 7. Size Exclusion Chromatography. Bigger molecules leaving the column first.

Hydrophobic Interaction Chromatography (HIC), Fig. 8, 9.

The HIC media exhibits amphiphilic properties, featuring both hydrophobic and hydrophilic regions, which facilitate the separation of proteins based on their surface hydrophobicity, or ability to repel water molecules. Target proteins and their associated aggregates typically possess distinct hydrophobic characteristics, and their removal via HIC contributes to further purification of the protein of interest. Moreover, the operating environment in HIC generally involves less harsh denaturing conditions compared to other chromatography techniques, thereby aiding in the preservation of the protein's native structure and functionality. While interactions between the resin and the hydrophobic regions of the

protein would be weak in pure water, applying a protein sample to HIC resin in a high ionic strength buffer enhances this interaction. Subsequently, reducing the ionic strength of the buffer facilitates the elution of proteins in order of decreasing hydrophobicity.

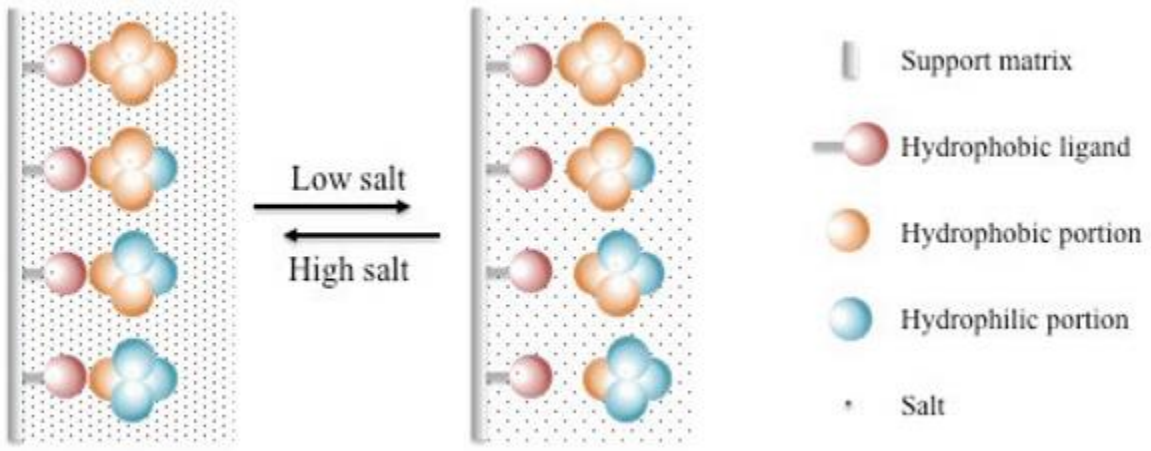


Figure 8. Hydrophobic Interaction Chromatography (Source: Chen et al, 2015)

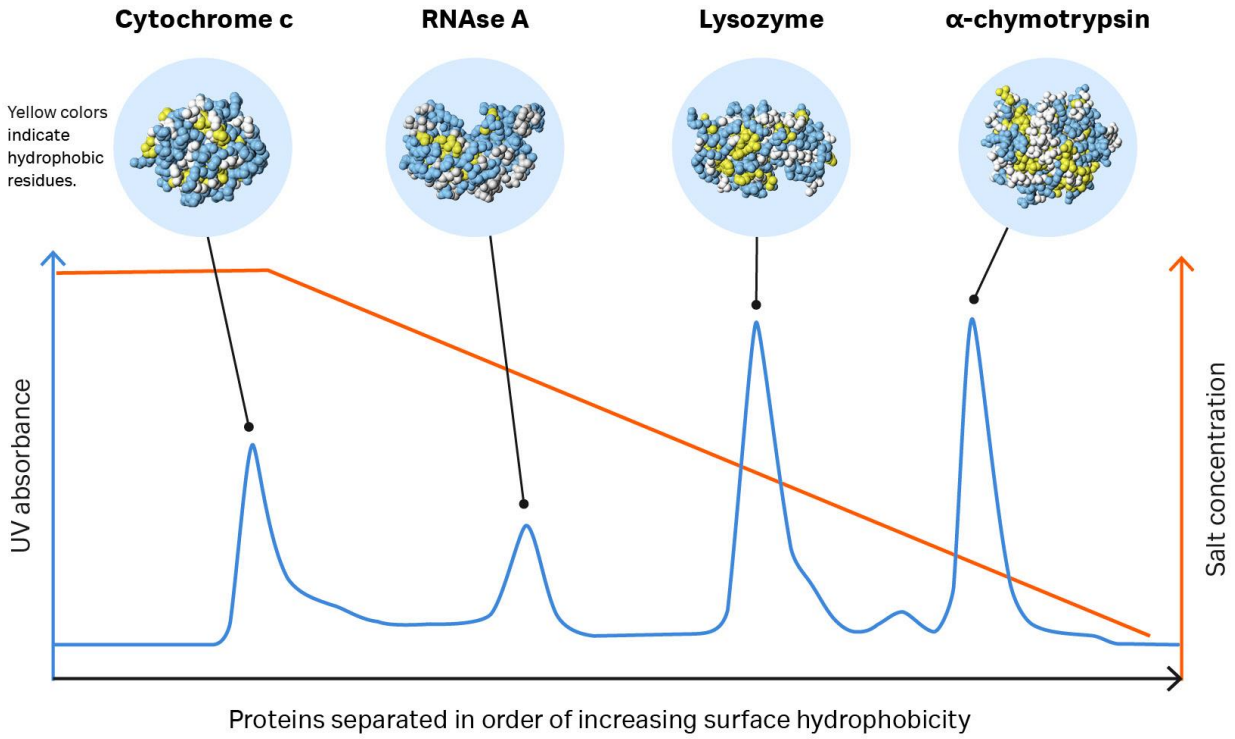


Figure 9Fig. 9. Proteins separate according to differences in their surface hydrophobicity (yellow indicates hydrophobic amino acid residues and blue indicates hydrophilic amino acid residues). Source: Cytiva(R)

Ion exchange chromatography, Fig. 10.

Ion exchange chromatography sorting compounds based on their ionic charge and characteristics. The choice of column depends on its charge type and intensity. Anion exchange resins, where the stationary phase ligands are positively charged, are utilized to retain and segregate negatively charged molecules (anions), whereas cation exchange resins, which, in turn, are negatively charged, are employed for positively charged molecules (cations).

Before separation commences, a buffer is passed through the column to balance out the opposing ions on the column's resin. When the sample is introduced, solute molecules exchange with the buffer ions as they vie for binding sites on the resin. The retention time for each solute relies on the strength of its charge. Compounds with weaker charges elute

first, followed by those with progressively stronger charges. Given the nature of this separation method, pH, buffer type, buffer concent

ration, and temperature all play crucial roles in governing the separation process.

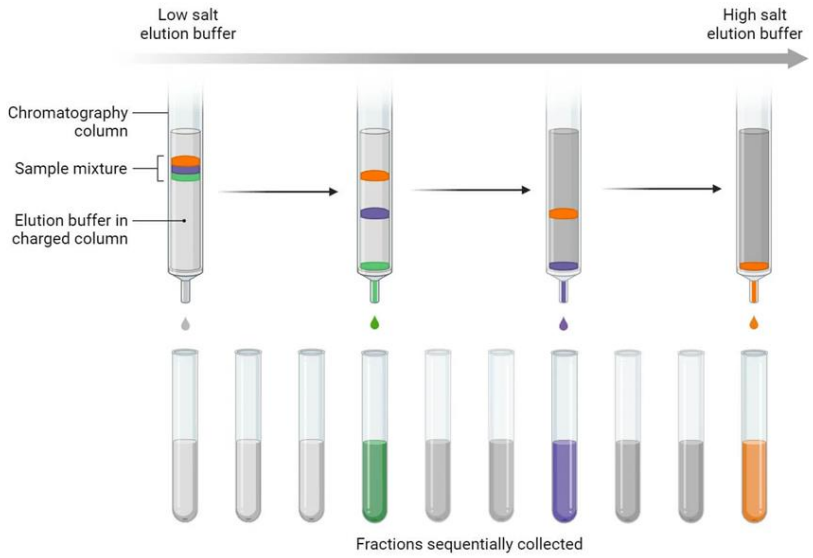


Figure 10. Ion exchange Chromatography principles

Affinity Chromatography

Affinity Chromatography relies on the specific molecular structure of compounds and often employs tailor-made resins for applications. These resins are equipped with ligands on their surfaces that are designed to bind selectively with the compounds targeted for separation. Typically, these ligands operate similarly to the interactions seen in antibody-antigen binding, resulting in a highly specific "lock and key" fit between the ligand and its target compound, often leading to a single peak while the rest of the sample remains unbound.

For instance, many membrane proteins are glycoproteins and therefore can be purified using lectin affinity chromatography. In this process, detergent-solubilized proteins are allowed to bind to a chromatography resin modified with covalently attached lectins. Proteins that do not bind to the lectin are washed away, and specifically bound glycoproteins can be eluted

by introducing a high concentration of sugar, which competes with the bound glycoproteins at the lectin binding site. However, some lectins have a strong affinity for the oligosaccharides of glycoproteins, making it challenging to displace them with sugars, and in such cases, the bound glycoproteins may need to be released by denaturing the lectin.

Another example of Affinity Chromatography is the His-tag or Strep-tag affinity process, where the surface of resin particles is covered with appropriate ligands: NTA (Nitrilotriacetic Acid) or IDA (Iminodoacetic acid) in the case of the His tag, and biotin in the case of the Strep tag. There exists a wide variety of affinity resins, and downstream processing scientists can choose which resin to use according to their specific requirements.

It's also worth noting that Affinity Chromatography is very often utilized as a capturing step or the first step of purification, due to its robustness and ease of implementation.

A specialized form of affinity chromatography known as Immunoaffinity chromatography utilizes the specific interaction between an antibody and its antigen (the target molecule the antibody selectively binds to purify the desired protein. This technique involves attaching an antibody to a solid support, such as a porous bead or a membrane, which then selectively captures the target molecule while allowing everything else to pass through. Elution of the target protein can be achieved by altering the pH or salinity of the system. The immobilized ligand may consist of an antibody (e.g., Immunoglobulin G) or a protein (e.g., Protein A). Because this method does not require the addition of a tag, it is suitable for purifying proteins sourced from natural origins.

HPLC vs FPLC techniques

High-performance liquid chromatography (HPLC), also known as high-pressure liquid chromatography, employs elevated pressure to expedite the passage of solutes through the column, thereby limiting diffusion and enhancing resolution. The most prevalent variant is "reversed phase" HPLC, wherein the column material exhibits hydrophobic properties. Proteins are eluted through a gradient of water and increasing concentrations of an organic solvent, such as acetonitrile, based on their hydrophobicity. HPLC purification often results

in denaturation of the purified proteins, rendering it unsuitable for proteins incapable of spontaneous refolding.

To address the limitations of HPLC, an alternative method employing lower pressure was developed, known as Fast protein liquid chromatography (FPLC). FPLC is a liquid chromatography technique commonly utilized for the analysis or purification of protein mixtures. Like other chromatography forms, separation occurs due to the differential affinities of various components for two materials: the "mobile phase and the stationary phase. In FPLC, the mobile phase consists of an aqueous buffer, with the buffer flow rate regulated by a positive-displacement pump, typically maintained at a constant rate. However, the buffer composition can be altered by drawing fluids from multiple external reservoirs in varying proportions. The stationary phase comprises resin beads, typically cross-linked agarose, packed into a cylindrical glass or plastic column. FPLC resins are available in diverse bead sizes and surface ligands, tailored to specific applications.

Developed and introduced to the market by Pharmacia in Sweden in 1982, FPLC initially bore the name "fast performance liquid chromatography" to distinguish it from HPLC. While FPLC is primarily employed for protein analysis, its versatility is enhanced by the wide array of available resins and buffers, enabling diverse applications. Unlike HPLC, FPLC operates under relatively low buffer pressure, typically below 5 bar, yet maintains a relatively high flow rate, typically ranging from 1 to 5 ml/min. FPLC is adaptable to various scales, accommodating the analysis of milligram-level mixtures in columns with volumes of 5 ml or less, as well as industrial-scale production, yielding kilograms of purified protein in columns with volumes spanning many liters. To familiarize with typical FPLC system components refer to Fig. 11.

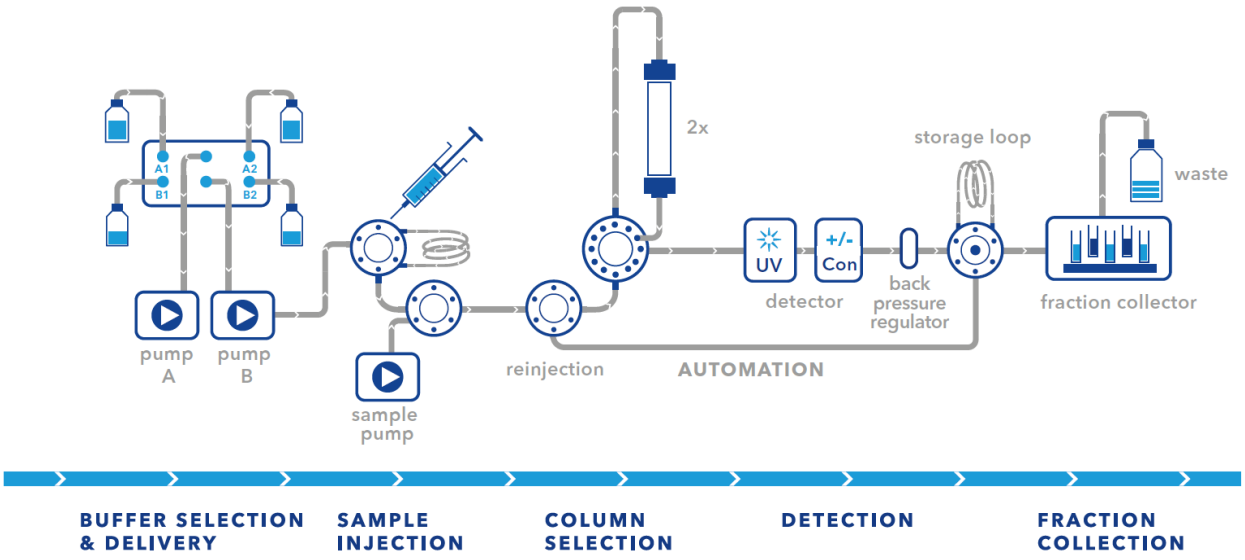


Figure 11. Typical schematics of FPLC System (Source: KNAUER Wissenschaftliche Geräte GmbH)

Formulation and Fill-Finish operations

After the successful purification of the desired protein, the final product usually undergoes several additional steps. First, appropriate testing of the product must be done to ensure the identity and purity of the protein. We will discuss a variety of analytical methods in the appropriate chapter. However, after confirming the product quality, a few final steps still need to be completed. These may include formulation, potential cold storage (freezing/thawing), and freeze-drying. Because most protein molecules are marginally stable due to their small free energy of stabilization (Dill 1990), stabilization steps are needed. Additionally, various degradation processes are inherent in most protein compositions. Aggregation, absorption by container walls, and chemical degradation are among the most common.

To mitigate the harm, incurred from these processes and ensure the high quality of the protein product, scientists and process engineers may need to add specific chemical components to the final product, determine the most stable solution composition, ascertain

the final product state (liquid, powder, etc.), select the primary packaging container, and establish the storage conditions. All these aspects fall under the purview of formulation science, which we will discuss in the appropriate chapter.

Upscaling of Protein Production and Commercialization for final product

Once the method of protein purification has been developed, it often needs to be scaled up, or expanded to produce a larger quantity of the product. The reasons for this may include:

- Growing demand for certain proteins (e.g., more enzymes for therapeutics or cosmetic products).
- Economic reasons: larger-scale production often results in a cheaper price per unit of goods.
- Consistency and Quality: Upscaling production allows for better control over production parameters, leading to improved consistency and quality of the final product. This is particularly important for proteins intended for therapeutic use, where batch-to-batch consistency is vital.
- Regulatory Requirements: Regulatory bodies often require large-scale production data and consistency studies to approve protein products for commercial use. Therefore, upscaling production is necessary to comply with regulatory requirements and obtain necessary approvals."
- Commercialization: For proteins developed for therapeutic or industrial purposes, upscaling is necessary to transition from laboratory-scale production to commercial manufacturing. This is a critical step in bringing a protein product to market.

We must say that commercialization of engineered proteins presents several significant challenges:

1. Production Scale-up: Transitioning from laboratory-scale production to large-scale commercial manufacturing can be complex and costly. Developing scalable

production processes that maintain product quality and consistency at a larger scale is a major challenge.

2. Cost-effectiveness: Commercial production of engineered proteins must be economically viable to compete in the market. This includes optimizing production processes to minimize costs while ensuring high yields and quality.
3. Regulatory Compliance: Engineered proteins intended for therapeutic or industrial use must meet stringent regulatory requirements for safety, efficacy, and quality. Obtaining regulatory approvals can be a lengthy and costly process.
4. Product Stability: Engineered proteins may be susceptible to degradation, aggregation, or other stability issues during production, storage, and transportation. Ensuring product stability throughout its lifecycle is crucial for commercial success.
5. Purification and Characterization: Purifying and characterizing engineered proteins to meet regulatory standards and ensure product quality can be challenging, especially for complex protein products.
6. Intellectual Property Protection: Securing intellectual property rights for engineered proteins and navigating potential patent disputes can be critical for protecting investments in research and development and maintaining a competitive advantage in the market.
7. Market Acceptance: Convincing customers of the benefits and value of engineered proteins compared to existing products or alternatives can be a significant challenge, particularly in industries with established products or practices.
8. Supply Chain Management: Managing the supply chain for engineered proteins, including sourcing raw materials, manufacturing, distribution, and quality control, requires careful coordination and oversight to ensure product integrity and meet customer demand.

Overall, addressing these challenges requires interdisciplinary collaboration, innovative technologies, and strategic planning to successfully commercialize engineered proteins and bring them to market.

Methodology and data collection

Protein Manufacturing Process development

Protein manufacturing process development includes all upstream (expression) and downstream (purification) steps described above. In the following sections, we will review the development of each step according to current industry practices, suggest some optimizations for the processes, and describe how we used it to collect our experimental data.

Quality system and Applied Analytics

A quality system in protein production encompasses the set of processes, procedures, and standards put in place to ensure that the manufactured proteins meet predefined quality criteria. It serves as a comprehensive framework that governs every stage of protein production, from upstream expression to downstream purification and beyond. By implementing a robust quality system, protein manufacturers can uphold product consistency, safety, and efficacy while adhering to regulatory requirements and industry best practices. This introduction will explore the key components and importance of quality systems in protein production.

1. **Quality Assurance (QA):**
    - **Definition:** QA refers to the planned and systematic activities implemented within the quality system to ensure that processes are designed, executed, and monitored effectively to produce protein products that meet predetermined quality standards.
    - **Responsibilities:**
        - Developing and implementing quality policies and procedures.

- Conducting audits and assessments to ensure compliance with quality standards and regulations.
- Providing training and support to personnel involved in protein production to maintain quality standards.
- Establishing systems for document control, change control, and deviation management.

- **Objective:** The primary objective of QA is to prevent defects and deviations in the production process and to ensure that protein products meet quality specifications consistently.

2. **Quality Control (QC):**
    - **Definition:** QC refers to the set of activities and procedures conducted during and after protein production to assess the quality of raw materials, intermediate products, and final protein products.
    - **Responsibilities:**
        - Performing analytical testing and assays to evaluate the purity, potency, and stability of protein products.
        - Monitoring and documenting critical process parameters to ensure consistency and reproducibility.
        - Conducting inspections and validations of equipment and facilities to ensure they meet quality standards.
        - Investigating deviations and non-conformances to identify root causes and implement corrective and preventive actions.
    - **Objective:** The primary objective of QC is to verify and validate that protein products meet predetermined quality specifications and regulatory requirements.
3. **Regulatory Compliance:**

- **Definition:** Regulatory compliance involves adhering to laws, regulations, and guidelines set forth by regulatory agencies, such as the FDA (Food and Drug Administration) or EMA (European Medicines Agency), to ensure the safety, efficacy, and quality of protein products.
- **Responsibilities:**
    - Interpreting and implementing regulatory requirements related to protein production processes, documentation, and quality standards.
    - Maintaining comprehensive documentation and records to demonstrate compliance with regulatory requirements.
    - Participating in regulatory inspections and audits and addressing any findings or observations.
    - Keeping abreast of changes in regulations and guidelines and implementing necessary updates to ensure ongoing compliance.
- **Objective:** The primary objective of regulatory compliance is to ensure that protein production processes adhere to applicable laws and regulations to protect public health and safety.

Overall, the integration of QA, QC, and regulatory aspects within the quality system of protein production ensures the consistent delivery of safe, high-quality protein products that meet regulatory requirements and customer expectations.

During protein production, various applied analytics and techniques are employed for Quality Control (QC) to assess the quality, purity, and consistency of the produced proteins. Some common applied analytics used for QC during protein production include:

1. **Chromatography Techniques:**
    - **High-Performance Liquid Chromatography (HPLC):** Used for the separation, identification, and quantification of proteins based on their size, charge, or affinity.

- **Size Exclusion Chromatography (SEC):** Determines the molecular weight distribution and oligomeric state of proteins.
- **Ion Exchange Chromatography (IEX):** Separates proteins based on their net charge.
- **Affinity Chromatography:** Utilizes specific ligands to selectively bind and purify proteins based on their affinity for the ligand.

2. **Electrophoresis Techniques:**

Invented by Arne Tiselius in 1930-s, this technique is based on a separation of different molecules in the polymeric matrix under force of an electrical fields. The bigger molecules moving slower than smaller ones. With

- **Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE):** Separates proteins based on their molecular weight.

In SDS-PAGE, proteins are denatured and coated with SDS, a strong anionic detergent. This treatment unfolds the proteins and gives them a negative charge proportional to their length, thereby neutralizing any intrinsic charge differences and allowing separation based primarily on size (Nowakowski et al., 2014).

There are two main types of SDS-PAGE gels: native and denatured.

Native SDS-PAGE Gel:

- In a native gel, proteins are separated without being denatured. They retain their native conformation, including any tertiary or quaternary structures they might have.
- Native gels are typically used when preserving the native structure of proteins is crucial, such as when studying protein-protein interactions or complex formation.

- Separation in native gels relies on factors other than size, such as charge and shape. Therefore, the resolution of native gels is often lower compared to denatured gels.
- Native gels are used less frequently compared to denatured gels because denatured SDS-PAGE provides more consistent results and is better suited for analyzing the molecular weight of proteins.

Denatured SDS-PAGE Gel:

- In denatured SDS-PAGE, proteins are treated with SDS and heated to a high temperature, typically around 95°C, to fully denature them and disrupt any non-covalent interactions, such as hydrogen bonds and hydrophobic interactions.
- Denatured gels are the most used type of SDS-PAGE gels. They provide high-resolution separation of proteins based on their molecular weight.
- Since all proteins are denatured and coated with SDS, the primary factor influencing separation is the molecular weight of the proteins. Therefore, denatured gels are ideal for determining the molecular weight of proteins and analyzing their purity.
- Denatured gels are extensively used in protein purification, analysis of protein samples, and studying protein expression levels.
- **Western Blotting (Immunoblotting):** Detects specific proteins using antibodies and is useful for protein identification and quantification.

3. **Mass Spectrometry (MS):**

- **Liquid Chromatography-Mass Spectrometry (LC-MS):** Identifies and quantifies proteins based on their mass-to-charge ratio.
- **Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-MS):** Analyzes intact proteins and peptides, providing information on their molecular weight and structure.

4. **Spectroscopic Techniques:**

    - **UV-Visible Spectroscopy:** Measures the absorbance of proteins at specific wavelengths to determine concentration and purity.
    - **Circular Dichroism (CD) Spectroscopy:** Provides information on protein secondary structure, folding, and stability.
    - **Fluorescence Spectroscopy:** Assesses protein folding, conformational changes, and interactions based on fluorescence properties.

5. **Biological Assays:**

    - **Enzyme-Linked Immunosorbent Assay (ELISA):** Detects and quantifies specific proteins using antibodies and enzymatic reactions.
    - **Bioactivity Assays:** Measure the functional activity of proteins, such as enzyme activity, receptor binding, or cell-based assays.

6. **Bioinformatics Tools:**

    - **Sequence Analysis:** Analyzes protein sequences for identity, similarity, and post-translational modifications.
    - **Structure Prediction:** Predicts protein structure and folding patterns using computational methods.

These applied analytics are essential for monitoring critical quality attributes, ensuring batch-to-batch consistency, and meeting regulatory requirements during protein production. They provide valuable insights into protein quality, purity, stability, and functionality, facilitating process optimization and product characterization.

Critical process parameters for establishing robust protein production pipeline.

Establishing a robust protein production pipeline involves identifying and controlling critical process parameters (CPPs) to ensure consistent and high-quality protein yields. These parameters may vary depending on the specific production system, and the protein being produced. However, here are some general critical process parameters commonly considered in protein production:

1. **Cell Line**: Selecting an appropriate host cell line (e.g., E. coli, yeast, mammalian cells) with high protein expression capability and desired post-translational modification machinery.
2. **Culture Medium**: Designing a suitable culture medium with optimal nutrient composition to support cell growth and protein expression.
3. **Temperature**: Controlling the temperature throughout the production process to maintain cell viability and protein stability.
4. **pH**: Monitoring and adjusting the pH of the culture medium to maintain optimal cell growth and protein folding conditions.
5. **Agitation and Aeration**: Providing adequate agitation and aeration to ensure uniform mixing of nutrients and oxygen distribution, which are crucial for cell growth and protein production.
6. **Induction Strategy**: If using an inducible expression system, optimizing the timing, concentration, and duration of inducer addition to maximize protein expression while minimizing cell stress.
7. **Cell Density**: Monitoring and controlling cell density to avoid overcrowding, which can lead to nutrient depletion and decreased protein yields, or excessive cell death.
8. **Harvest Time**: Determining the optimal time to harvest cells or supernatant to maximize protein yield while minimizing degradation or proteolysis.

9. **Purification Strategy**: Developing efficient purification processes tailored to the specific properties of the target protein, including chromatography conditions, buffer compositions, and filtration methods.
10. **Quality Control Assays**: Implementing robust assays to monitor protein quality, including purity, integrity, and activity, throughout the production process.
11. **Environmental Conditions**: Maintaining consistent environmental conditions such as humidity, light exposure, and sterility to minimize variability in protein production.
12. **Upstream and Downstream Integration**: Ensuring seamless integration between upstream (cell culture) and downstream (purification) processes to optimize overall efficiency and yield.
13. **Scale-up Considerations**: Anticipating and addressing potential challenges associated with scaling up production from laboratory to industrial scale, including equipment limitations, process intensification, and cost considerations.

Besides those general considerations we also will review critical parameters for each step of the production pipeline:

Correct host organism and appropriate cloning method.

Choosing the correct cloning method for engineering protein production depends on various factors, including the host organism, desired protein characteristics, expression levels, and downstream processing considerations. The most crucial factor is the characteristics of the final product, such as its type, functionality, and yield. As mentioned earlier, if a fully functional, correctly folded, and post-translationally modified protein is desired, mammalian cells serve as the preferred host and expression system. However, recent progress in developing bacterial, yeast, and plant expression systems has enabled the production of high yields of functional products therein. Nowadays, selecting the correct host and cloning methods has largely become a business decision, focusing

mostly on cost, speed, and development expenses. Let's review the selection of hosts and cloning methods for the main modern expression systems, as different hosts require different cloning methods.

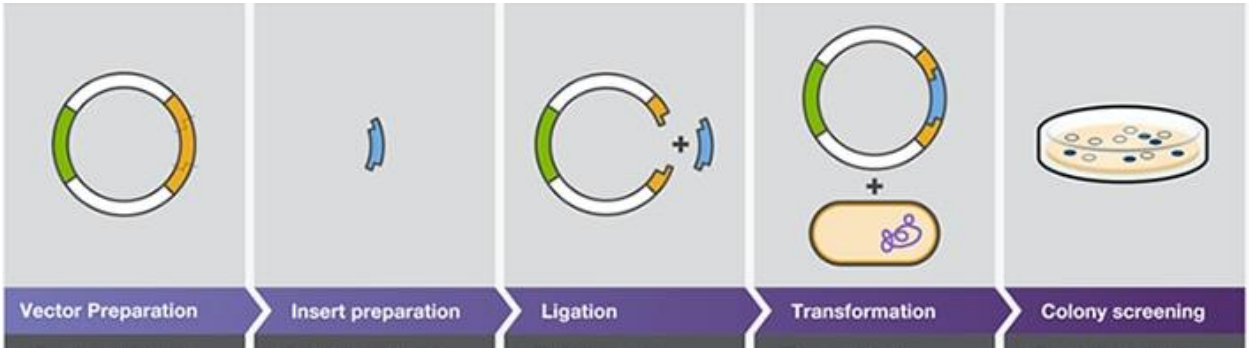


Figure 12. Flow chart of molecular cloning steps (Thermo Fisher Scientific®).

In general, molecular cloning is a technique that allows the assembly of recombinant DNA molecules and their replication within a host organism (Watson, 2007). A generic workflow for these experiments typically includes the following steps (see Fig. 12):

1. Vector Preparation – Selecting a DNA molecule (usually viral or bacterial) that will carry the desired DNA coding fragment into a host organism.
2. Insert Preparation – Isolating or synthesizing a specific DNA sequence that codes for the desired protein of interest, and then preparing it for insertion into a vector.
3. Ligation – Combining the vector and insert into one construct.
4. Transformation – Delivering the vector with the insert into the host organism.
5. Colony Screening – Selecting the cells (a.k.a. clones) that have a functional construct. That is, the vector is incorporated into the host cell structures and expresses the protein of interest.

Below, we describe cloning methods specific to each expression system:

For **plant cells** Agrobacterium mediated transformation or particle bombardment are predominant methods to deliver foreign DNA into the cells. These methods involve transferring the desired DNA fragment into Agrobacterium, which then infects plant cells, integrating the DNA into the plant genome, or coating DNA onto microscopic gold particles, which are propelled into plant cells.

For **bacterial cells** a variety of methods to create a construct is used: ligation independent cloning, TA cloning, site-directed mutagenesis, Gibson assembly, and homologous recombination are among the most used methods. For the transformation heat shock and electroporation are used.

For **yeast cells** many of bacterial methods can successfully be used along with electroporation as a primary transformation method.

For **insect cells** *Baculovirus* based expression vectors are predominantly used. These vectors typically contain strong viral promoters, such as the polyhedrin or p10 promoters, to drive high-level expression of the gene of interest, while calcium phosphate precipitation, electroporation, or lipid-mediated transfection are the most often used methods of the delivery of DNA construct into the cell.

For **mammalian cells** pcDNA vectors, lentiviral based and adenoviral based viruses are among the most popular. Using adeno and lentiviral systems based on their possibility to successfully integrate into host genome, provide high level expression of proteins, and low immunogenicity and genotoxicity (reducing the likelihood of insertional mutagenesis). Ease of use and efficient transduction of the vectors also making them popular among research community.

Different hosts sometimes require completely different cloning techniques that must correspond to the cellular machinery of the host. However, there are common critical parameters that need to be considered, with the most important being:

1. Vector Compatibility: It is imperative to choose a cloning method compatible with the vector system intended for protein expression, such as viral vectors or bacterial artificial chromosomes (BACs).

2. Expression System: It is also necessary consider the expression system to be used, including constitutive or inducible promoters and secretion signals. Certain cloning methods are better suited for specific expression systems.

3. Protein Size and Complexity: Assessing the size and complexity of the protein of interest is another necessary factor to consider. Certain cloning methods may be more appropriate for cloning large or complex proteins due to their efficiency and fidelity.

4. Post-translational Modifications: If the protein requires specific post-translational modifications (e.g., glycosylation, phosphorylation), selection a cloning method compatible with the host organism's post-translational modification machinery may play a crucial role for protein production.

5. Restriction Sites and Assembly Strategy: Probably most critical step in vector-insert design is evaluation of restriction sites available and choosing a cloning method that allows for seamless assembly of the gene of interest into the expression vector. Common assembly strategies include restriction enzyme-based cloning, Gibson assembly, and ligase-independent cloning (LIC).

6. Compatibility with Downstream Processing: Finally, cloning method have to be compatible with downstream processing steps such as protein purification and characterization. Some methods may introduce tags or fusion partners that facilitate purification, while others may require additional steps for tag removal.

By carefully considering these factors, one can select the most appropriate cloning method to efficiently and effectively engineer protein production for specific research or industrial applications. Additionally, consulting recent literature and seeking advice from experienced researchers in the field can help guide the decision-making process.

Correct extraction/purification method

Choosing the correct extraction and purification methods is both an art and a science, involving a delicate balance between scientific principles and business considerations. On the scientific side, it requires a deep understanding of the biochemical properties of the target protein, the intricacies of various purification techniques, and the ability to troubleshoot and optimize protocols. On the business side, decisions must take into account cost-effectiveness, scalability, and regulatory compliance. Unless the protein of interest is secreted into the media surrounded expression cells it is necessary to break the cell walls to make the protein available for the following downstream process steps. Historically, mechanical disruption and detergent-based cell lysis were most widely used methods for that. With currently developed technology there are many optimized equipment and formulations to complete the task.

Extraction proteins from higher plants, however, can pose unique challenges due to specific of plant anatomy. Very often, protein of interest is accumulating either in specific organelles (for example chloroplasts) or periplasmic area. In those cases, combination of proper mechanical disruption and correctly chosen extraction buffer is necessary.

Very often, host cells contain proteolytic enzymes, oxidating agents who in turn can cause severe protein degradation at the very early stages of downstream process. To prevent that, protease inhibitors (for example Thermo Fisher Scientific® protease inhibitor cocktails) and antioxidants can be added to the lysis/extraction buffer. Lowering processing temperature also can help to prevent protein degradation. That is especially important in the case of sensitive and unstable proteins.

After cell wall disruption, correct clarification method should be chosen. Most common methods as we mentioned in previous chapters are centrifugation and filtration. Choosing a correct method depends on the scale and desirable purity of the product entering a purification workflow.

For example, for mammalian cells (assuming they are grown in bioreactor) the typical process can include cell lysis, centrifugation, decanting of the supernatant, filtration

through 0.45- or 0.2-micron fiberglass filters to get rid of residual particulate materials. Then this material can enter chromatographic workflow.

Similar process can take place for bacterial, yest and insect cells. In case of the higher plant cells more steps maybe needed. First, plant material should undergo thorough homogenization, then several steps for preliminary filtration needs to be done to separate liquid part of the biomass from the fibers and cells debris. For that, filtration through dolomite powder or ceramic-based filters are most common methods applied by the industry.

Next, and probably most expensive step for protein production is the purification and many difficult decisions must be done at this stage. Choosing correct purification scheme can "make or break" the entire protein production pipeline.

Here are **critical factors** need to be considered:

Necessity, number, and characteristics of the intermediate filtration steps. Do we need any ultrafiltration steps to get rid of higher/lower molecular weight contaminants? Do we need buffer exchange between steps? Any specific filters needed (catching endotoxin, heavy metals, specific immunogens)?

Biochemical characteristics of the protein of interest: pI(isoelectric point), molecular weight, stability index, hydrophobicity regions, disordered regions, specifically active domains, tertiary and quarterly structure (is it monomer, multimer, or complex protein).

Based on the above information specific chromatographic steps need to be chosen. Typical purification contains two steps capturing and polishing chromatography.

In capturing chromatography, a stationary phase (often a resin or gel matrix) is packed into a column, and the target proteins in the sample are selectively captured onto this stationary phase while other components in the mixture pass through the column without binding. The selectivity of capturing chromatography is achieved through specific interactions between the target biomolecules and the ligands or functional groups present on the surface of the stationary phase.

Once the target biomolecules are captured on the stationary phase, they can be eluted from the column using various methods such as changing the pH, ionic strength, or composition of the mobile phase. This elution step allows for the separation and collection of the purified biomolecules for further analysis or downstream applications.

Capturing chromatography is often the initial step in a multi-step purification process, followed by additional chromatographic steps or techniques to achieve higher purity and yield of the target biomolecule. Common types of capturing chromatography include affinity chromatography, ion exchange chromatography, and hydrophobic interaction chromatography, each of which exploits different molecular interactions for selective biomolecule capture.

In protein chromatography, the "polishing step" refers to the final stage of a purification process aimed at achieving high purity of the target protein. After the initial capturing step, which isolates the target protein from the crude sample, the polishing step further purifies the protein and removes any remaining contaminants or impurities to meet the desired level of purity for downstream applications. Choosing capturing step usually depends on the most prominent protein features. For example, for capturing antibodies protein A or protein G resin is a natural choice, for his-tagged proteins Ni or Co chelating resigns capture most of the protein and so on. Sometimes researchers need to develop very specific resin to capture a unique protein. A SARS-CoV-2 Spike protein RBD protein affinity resign, developed by Repligen® and Navigo Proteins can be a good example of that.

Typical methods used in the polishing step of protein chromatography include Size Exclusion, Ion Affinity, and Multimodal chromatography. These methods can be used individually or in combination, depending on the specific properties of the target protein and the contaminants present in the sample. Researchers have to pay attention to specific selectivity of the resign and choose the methods focused on the desired level of purity, the characteristics of the sample, and the downstream applications of the purified protein.

Correct final formulation (user end)

After purification, proteins typically exist in a concentrated solution. It is then should be sent to the analytical laboratory for further analysis and can be utilized for research and development purposes, marking the end of its production cycle.

This step is often overlooked by many researchers. However, if we aim to develop a protein production pipeline to produce products for consumer use, the form in which the protein is delivered becomes crucial. This is where the formulation process begins.

Formulation in protein production involves transforming a purified protein mixture to possess properties desired by end-users. For therapeutic proteins, these properties may include stability, preservation of activity, and ease of administration. For food additives, considerations may include solubility, absence (or presence) of specific taste, and oxidative stability, among others.

The formulation process entails adding various components to the purified protein product to alter its physical state and chemical properties, making it suitable for use by the customer.

Most final formulations nowadays exist in liquid form; however, solid and aerosol formulations also exist. Below, we describe each of these formulation types based on the state of the product and common methods for their development. It's important to note that the pharmaceutical industry has its own formulation classifications based on delivery methods and routes of administration.

Liquid formulations make a significant proportion of total protein products, particularly for therapeutic biologics. The advantages of liquid formulations are evident: they require fewer production steps and less equipment, making them less labor consuming. For therapeutic products, they can be administered in various ways, are easy to inject, and can remain stable for long periods. However, concentration variations, stability issues, and transportation challenges can pose certain difficulties for liquid protein formulations. For example, some therapeutics require a consistent cold chain and specific concentration limits for storage capacity.

Solid formulations are necessary, for example, for topical application and mixing with food products. These formulations can be achieved by freeze-drying (lyophilizing) a liquid formulation, which can conserve proper tertiary structure and functions of the proteins and improve overall stability of the molecule. Below, we will describe each step of development for this process.

Aerosol formulation occupies a unique niche among all formulations and is mostly applied to specific drugs requiring certain routes of administration. It begins with the development of a stable form of the drug substance, which is then mixed with aerosol-forming chemicals.

A general guideline for developing liquid formulations (Moreira and Sarraguça, 2020) includes defining the objectives (i.e., application, dosage form, and desired properties), selecting suitable excipients, optimizing formulation composition, and assessing formulation stability. Developing a new liquid formulation usually starts with screening a proper buffer composition. Buffers are essential for maintaining pH stability, controlling the ionization state of protein functional groups, preventing aggregation, preserving enzyme activity, and ensuring experimental reproducibility. Choosing the appropriate buffer system based on the desired pH range, buffer capacity, and compatibility with the protein of interest is critical for achieving optimal protein stability in formulation development.

The next step is excipient screening. The most common strategy is based on choosing a library of excipients that have been commonly used in protein formulations. These excipients typically include stabilizers, surfactants, bulking agents, tonicity agents, and preservatives. Below is an example of the development schematic for formulating a single biologic (non-specified Antibody) (Fig. 13).

First, the purified protein in the elution buffer obtained in the polishing step is dialyzed into different buffers within a certain pH range (usually plus or minus 1 from the isoelectric point of the molecule), then initial stability is assessed by dynamic light scattering (DLS), differential scanning calorimetry (DSC), or size-exclusion

chromatography (SEC). The most stable buffer/pH combination is then chosen, and excipient screening is performed.

The typical excipient screening technique workflow includes screening for one type of component per step, such as testing different salt concentrations or adding different sugars to the formulation and testing it. However, if the testing instrument capability allows it, multiple conditions can be screened simultaneously. For example, if the DLS instrument has a 96-well plate capacity, up to 32 different buffer/salt/sugar combinations of excipients (each condition in triplicate) can be assessed. This high-throughput screening (HTS) saves a great deal of time for researchers. These screening methods are usually combined with the Design of Experiments method (DoE). The DoE approach is a systematic and efficient method for planning, conducting, and analyzing experiments to understand the relationship between input variables (factors) and output responses. It enables researchers to gather maximum information with minimum experimentation, making it a powerful tool in various fields including engineering, science, and manufacturing.

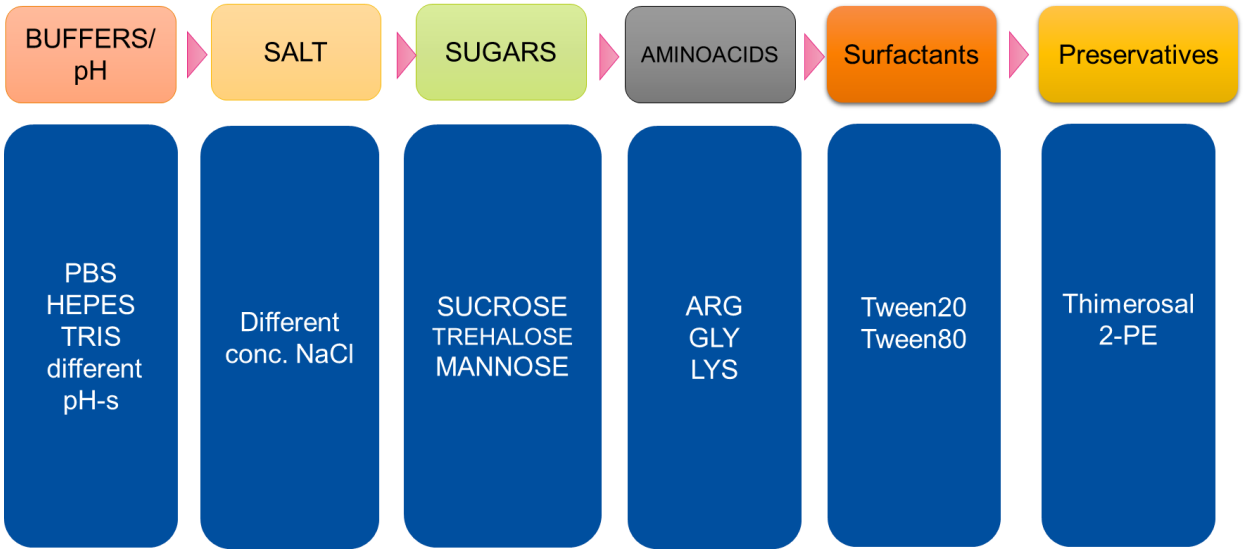


Figure 13. Workflow for formulation development of a non-specified Antibody (Author's work)

Correct analytics and QC

Each time protein production commences, it is imperative to ensure the accuracy of each production step's outcome. This not only guarantees product quality and readiness for subsequent production stages but also facilitates process enhancement by researchers, aiming for increased efficiency and robustness. For instance, if a higher level of product aggregation is observed during upscale purification compared to smaller pilot runs, adjustments can be made to elution parameters, inclusion of excipients in the buffer to prevent aggregation, or even a complete overhaul of purification strategies. Accurate analytics play a pivotal role in identifying issues during protein production, aiding in the development of stable products, and serving as the foundation for the final product's release to customers. Key inquiries posed to analytics/QC specialists typically revolve around the quantity, functionality, and stability of compounds throughout production cycles. Modern analytical departments deploy a diverse array of methods and instruments, many of which, such as SDS PAGE and HPLC, have been detailed in the QC section of this chapter. The standard workflow for obtaining samples for in-process and release testing is illustrated in Figure 14. In this workflow, samples for in-process testing are typically collected after each major production step, while stability and release testing are usually conducted for the final product. The data from the latter tests are crucial for generating a Certificate of Analysis/Compliance, which forms the cornerstone of release documentation.

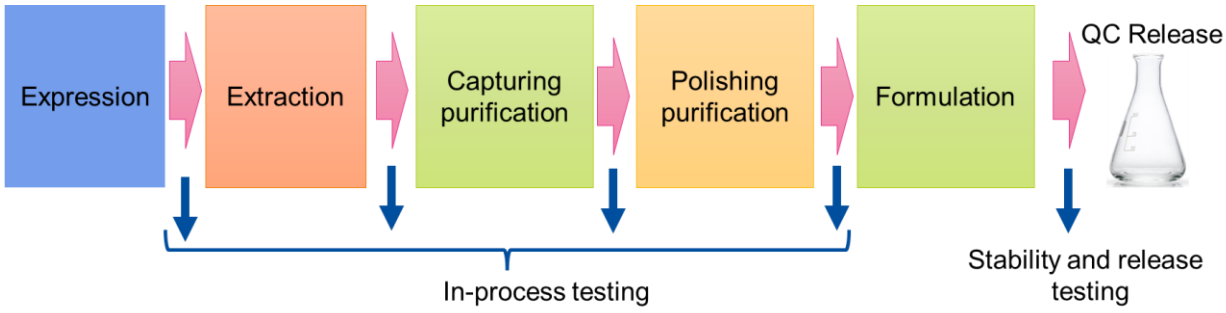


Figure 14. Typical Testing workflow for Protein production pipeline. (Author's work)

Summary of the Protein Manufacturing process development methods

The development of protein production methods is a multifaceted and intricate process that demands a systematic approach and the collaboration of a multidisciplinary team of specialists. Numerous critical parameters must be carefully monitored and, if needed, adjusted throughout this process. These parameters include the correct cloning method, selection of the appropriate host organism, identification of the optimal purification method, formulation of the final product, implementation of precise analytics, and rigorous quality control measures.

1. **Correct Cloning Method**: This parameter refers to the technique used to introduce the gene of interest into the host organism's genome or expression vector. It is crucial to select a cloning method that ensures accurate replication and expression of the desired protein.
2. **Correct Host**: The choice of host organism plays a significant role in protein production. Factors such as protein folding, post-translational modifications, and scalability can vary depending on the host. Selecting the correct host ensures efficient expression and yields of the target protein.
3. **Correct Purification Method**: After protein expression, purification is essential to isolate the target protein from cellular components and contaminants. The purification method must be chosen carefully to achieve high purity and activity of the final product while minimizing losses.
4. **Correct Final Formulation**: Formulation involves optimizing the composition of the purified protein product to meet the desired characteristics for its intended use. Factors such as stability, solubility, and compatibility with delivery methods must be considered to develop a formulation that maintains the protein's integrity and functionality.
5. **Correct Analytics and QC**: Accurate analytics and quality control are indispensable for assessing the quality, quantity, and functionality of the protein throughout the

production process. Various analytical techniques, such as SDS-PAGE, HPLC, and ELISA, are employed to monitor critical parameters and ensure product consistency and compliance with specifications.

By meticulously managing these critical process parameters, researchers can enhance the efficiency, reproducibility, and success of protein production methods, ultimately facilitating the development of novel therapeutics, diagnostics, and industrial enzymes.

Results

Case study 1

Developing protein production pipeline for structural genomic projects

Structural genomics is a field of molecular biology that aims to determine the three-dimensional structures of all proteins and RNA encoded by the genome of an organism. The primary goal of this discipline is to provide a comprehensive understanding of the structure, function, and interaction of biological macromolecules at a genomic scale.

Structural genomics has numerous applications in drug discovery, protein engineering, and understanding the molecular basis of diseases. By providing detailed structural information on a large scale, structural genomics accelerates the process of drug target identification, rational drug design, and understanding the mechanisms of drug action. It also contributes to fundamental understanding of biology by revealing the intricate architecture of proteins and their functional significance.

The main objective of SSGCID is to utilize cutting-edge high-throughput (HTP) techniques and methodologies to elucidate the three-dimensional structures of proteins and other molecules crucial to the biology of human pathogens, or molecules involved in interactions between hosts and pathogens (Myler et al., 2009). These pathogens encompass those classified under NIAID Category A-C agents, as well as emerging and re-emerging infectious disease organisms, on an annual basis for a span of five years. Additionally, the program provides backing for research endeavors employing experimental methods aimed at enhancing the characterization of protein targets' molecular mechanisms and gaining deeper understanding of their functional significance.

Overview of the SSGCID pipeline

The SSGCID Target Selection Team employs a comprehensive strategy involving bioinformatic and manual filters to select proteins from representative genomes of various bacterial, eukaryotic, and viral genera. This selection process aims to identify proteins with potential as drug targets or roles in cell growth, pathogenesis, or drug resistance, while filtering out proteins predicted to be challenging for soluble expression and crystallization. Each year, the team aims to select around 500 targets, with an additional 50-100 chosen based on community requests.

The selection process involves multiple batches, with each batch focusing on different criteria. Batch01 targeted potential drug targets in three bacterial species, while subsequent batches incorporated orthologues of actively pursued bacterial drug targets and expanded to include additional bacterial and eukaryotic species. Notably, Batch04 diverged from traditional structural genomics efforts by selecting RNA riboswitch elements from bacterial species (Myler et al., 2009).

After target selection, genes are cloned and expressed, initially in bacterial expression vectors. Targets failing to produce soluble protein undergo further screening using cell-free expression systems or synthetic gene construction. Additional tiers of screening are planned to further optimize protein production and potentially explore alternative expression systems.

Successful protein expression leads to purification at various protein production facilities. Crystallization screening follows, with high-throughput methods employed to identify crystallization conditions. Data collection and structure determination are attempted for crystals with suitable diffraction properties, with molecular replacement techniques commonly used for structure solution. For proteins resistant to crystallization, NMR spectroscopy is employed for structure determination.

Overall, the SSGCID employs a systematic approach from target selection to structure determination, utilizing a combination of traditional and innovative methods to overcome

challenges in protein expression and structure determination, ultimately contributing valuable insights into protein function and potential drug targets.

Example 1 Expression and purification of Mycobacterium transcription termination factor.

The bacterial Rho factor is a ring-shaped genome-wide transcription termination and R-loop dissociation protein. Rho is essential in many species, including in *Mycobacterium tuberculosis* where *rho* gene inactivation leads to rapid death, which makes it an attractive target for anti-TB drug discovery. However, expression and purification of this target is difficult task because first 200 amino acids in the sequence contribute to so called "disorganized region" which makes the full-length protein tend to aggregate. At the same time active center of the protein is located between AA 255 and 602 (Fig.15). Therefore, initial expression experiment was proposed to make 6 clones – from the full length to differently truncated constructs.

Preliminary experiments showed some errors including misalignment with proper restriction sites, absence of stop codons, frame-shift mutations, and presence of additional open reading frames (ORFs) beyond amino acid 602.

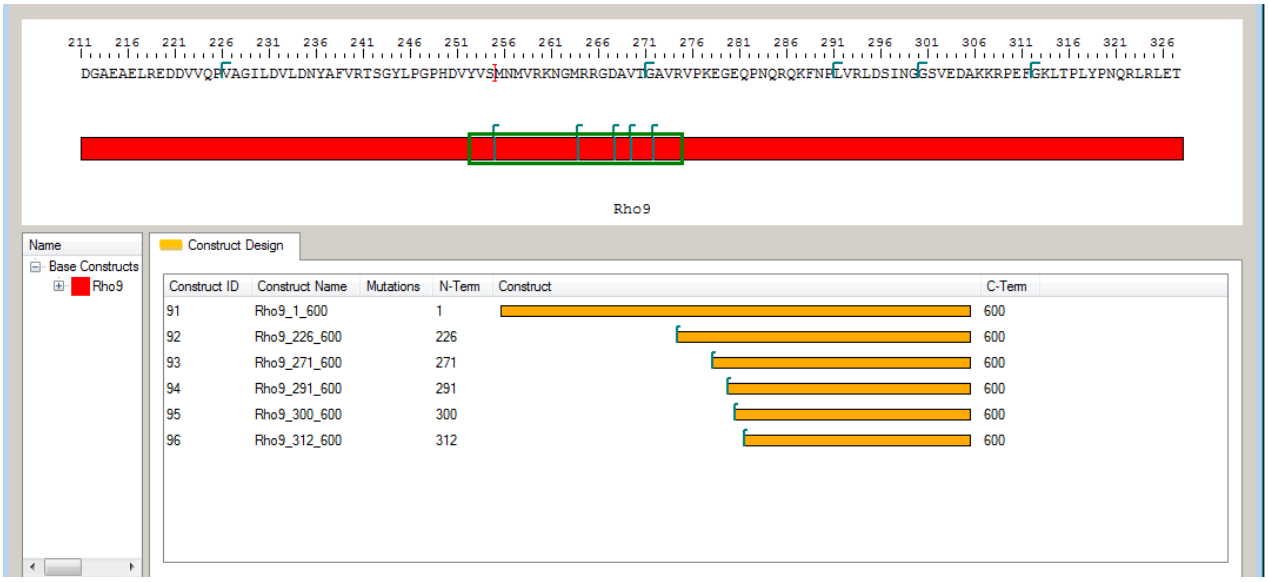


Figure 15. Rho9 clone's design

We ended up with the following study plan:

1. Design and Synthesis of New Infusion Primers:

- Development of new primers tailored to address the identified errors and facilitate precise amplification of the target sequences.

2. PCR Amplification of Clones 1-6:
    - Implementation of polymerase chain reaction (PCR) amplification using the newly designed primers to correct errors in clones 1-6.
    - Optimization of PCR conditions to ensure efficient amplification of the target sequences while minimizing errors.
3. Quality Control Assessment:
    - Rigorous QC evaluation of the PCR products to verify the successful correction of errors and ensure the integrity of the amplified sequences.
    - Utilization of techniques such as gel electrophoresis and sequencing to validate the accuracy and fidelity of the corrected clones.
4. Transformation into Expression Host:
    - Introduction of the corrected clones into the designated expression host system using appropriate transformation techniques.
    - Optimization of transformation protocols to enhance efficiency and ensure stable integration of the corrected sequences into the host genome.
5. Expression and IMAC Testing:
    - Induction of protein expression in the transformed host cells under optimized conditions.
    - Purification of the expressed protein using immobilized metal affinity chromatography (IMAC) to assess the efficiency and yield of protein production.
    - Comprehensive analysis of the purified protein samples through techniques such as SDS-PAGE, Western blotting, and functional assays to evaluate their integrity, purity, and biological activity.

6. Data Analysis and Interpretation:
    - Compilation and analysis of experimental data to assess the success of the corrective measures in eliminating cloning errors and improving protein expression.
    - Identification of any remaining challenges or areas for further optimization.
    - Interpretation of the results in the context of the original research objectives and implications for future studies.
7. Documentation and Reporting:
    - Documentation of all experimental procedures, observations, and results in a detailed and organized manner.
    - Preparation of comprehensive reports summarizing the study findings, including recommendations for future experiments or modifications to the experimental approach.

By following this structured study plan, we aim to address the previous cloning errors effectively and optimize protein expression, thus advancing our understanding of the target protein and its potential applications.

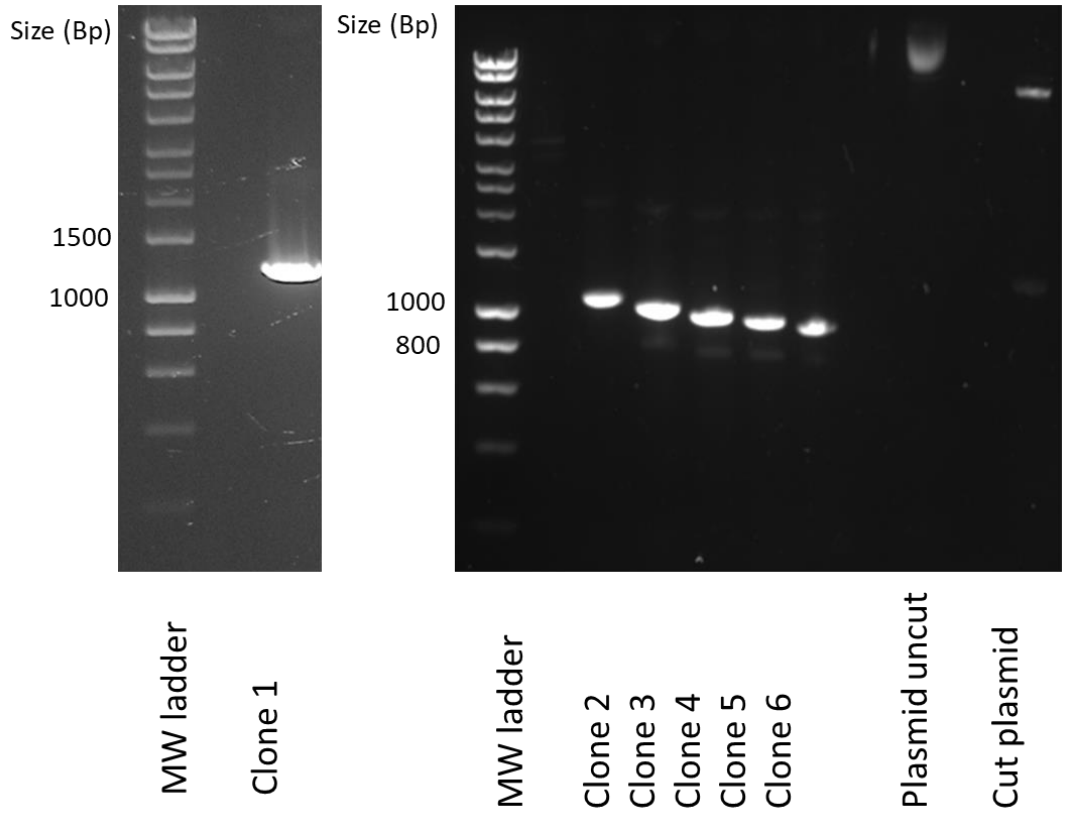
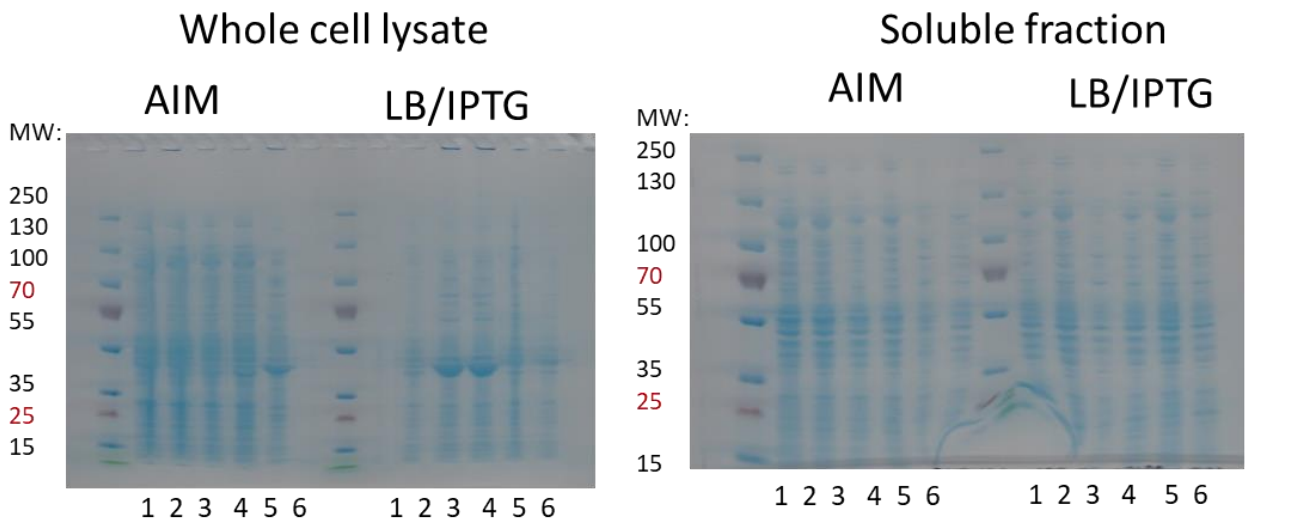


Figure 16. Amplification results of clones 1-6 of Rho9.

In the Figure 16 we can see all 6 clones amplified, purified and ready to clone into appropriate host. This time it was Rosetta E. coli strain. This strain was used because it is designed to enhance the expression of eukaryotic proteins that contain codons rarely used in E. coli. After infusion cloning and transformation Expression testing was completed (Fig. 17).



Legend:

1 – clone/construct 1, colony 1
2-clone /construct 1, colony 2
3- clone/construct 2, colony 1
4-clone/construct 2, colony 1
5- clone/construct 3, colony 1
6- clone/construct 3, colony 2

AIM – expressed using Autoinduction media
LB/IPTG - expressed using IPTG induction

Figure 17. Expression testing of first three clones of Rho9

All three clones expressing the desired construct, but the appropriate band in the soluble fraction was much smaller compared to whole cell lysate which indicate solubility issues while expressing the protein of interest. To solve that problem, temperature conditions during induction were optimized and below we can see Western blot of soluble fractions of clones 1,2, and 3 for autoinduction media and IPTG induction experiment respectively (Fig. 18, 19).

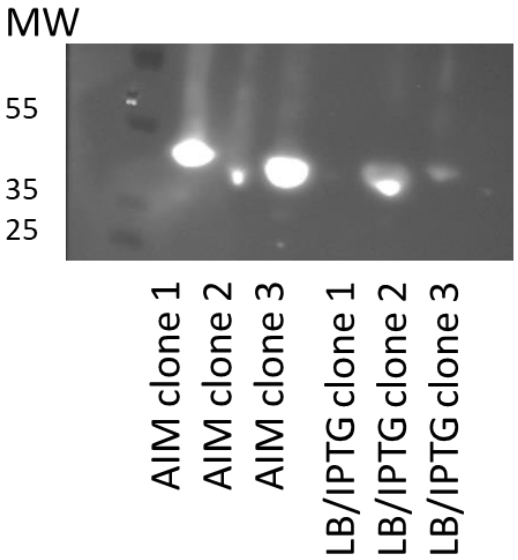


Figure 18. Western blot of soluble fractions for Expression testing of first three clones of Rho9 after optimization.

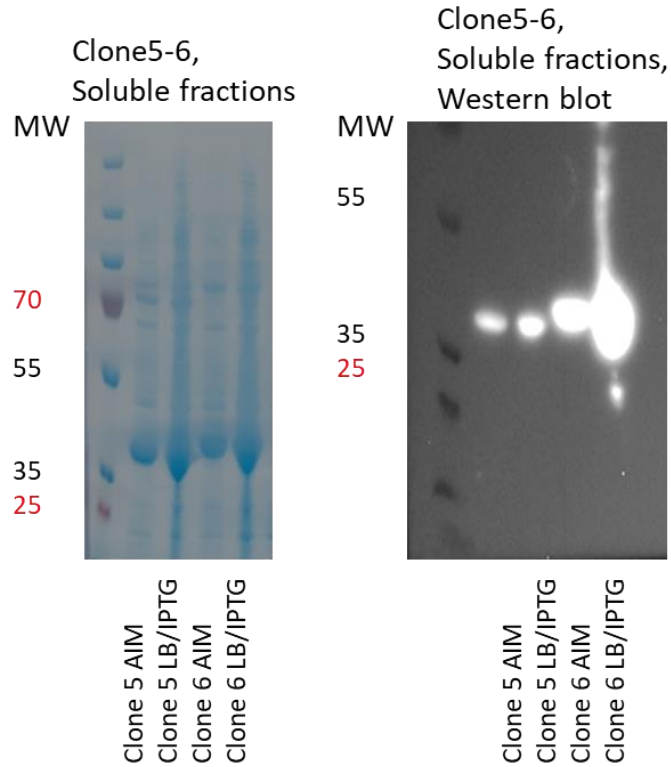


Figure 19. SDS PAGE and Western Blot of fractions 5-6 after optimization

Example 2 Expression and purification of SARS-CoV nsp 10 and nsp 14.

The RNA-synthesizing mechanism of SARS-CoV relies on 16 non-structural proteins, including nsp10, which interacts with nsp14 and nsp16 to enhance their enzymatic activities. By pinpointing key residues on nsp10's surface involved in its interaction with nsp14, researchers disrupted this interaction, hindering nsp10's activation of nsp14's exoribonuclease activity. This interaction region on nsp10 also regulates nsp16's 2′-O-methyltransferase activity. Crucially, mutations that blocked the nsp10-nsp14 interaction impaired SARS-CoV replication, highlighting nsp10's central role in orchestrating viral replicase function and suggesting potential therapeutic targets against SARS-CoV and related coronaviruses (Eckerle et al., 2010), Fig. 20.

Figure 20. SARS-CoV genome organization and ORF 1a/b polyprotein expression.

The genome is a 29.7-kb positive-sense RNA molecule that is capped (dark circle) and polyadenylated. Genes are indicated for the replicase (ORF 1a and ORF 1b; white), structural proteins [Spike (S), Envelope (E), membrane (M), and nucleocapsid (N) proteins; black], and accessory proteins (light gray). ORF 1b is accessed by ribosomal frameshift in the nsp12 coding sequence. The ORF 1a/b polyprotein is translated directly from input genome RNA and processed into 16 mature non-specific proteins by two virus-encoded proteinases (gray). (Eckerle et al., 2010)

In the protein production plan, we aimed to express two key RNA-synthesizing proteins, nsp10 and nsp14 from SARS-CoV, in Escherichia coli. To achieve this, we first genetically modified nsp10 with an N-terminal GST tag and cloned it into the pGEX-6p-1 vector. Similarly, nsp14 was tagged with a C-terminal His tag and inserted into the pRSFDuet-1 vector. Following this, we transformed these constructs into Rosetta strain E. coli cells. The transformed cells were then plated on specific antibiotic-containing plates; nsp10-expressing cells were plated on ampicillin (Amp) plates, nsp14-expressing cells on

kanamycin (Kan) plates, and cells expressing both proteins on plates containing both Amp and Kan.

After confirming successful transformations through colony growth, we selected three colonies from each plate and cultured them in LB media. Subsequently, we conducted small-scale expression and solubility tests for all clones obtained from each transformation. Remarkably, the results revealed consistent expression levels and patterns across all clones, indicating uniformity in protein expression and solubility among different colonies. This standardized procedure ensures reliability and reproducibility for further experiments aimed at characterizing and understanding the functions of nsp10 and nsp14 in the context of SARS-CoV replication.

In the process of working with bacteria to express and purify proteins, we often encounter various difficulties. When we found that the BL21 E. coli strain wasn't expressing enough protein, we tried using a different strain called LOBSTR. Thankfully, this change helped us get better results.

But as often happens in science, we hit another problem. When we looked closely at the mass spectrometry results, we found some unexpected contaminants in the solution, which turned out to be DnaK chaperones (Table 1). We had to act quickly to make sure our experiment's data was still accurate.

To fix this issue, we switched to using a different strain of E. coli called BL21(DE3) ΔdnaK. This strain didn't have the DnaK chaperones, so it helped us get rid of the contamination problem. This adjustment not only fixed our experiment but also showed how important it is to be flexible and problem-solve in science.

Even though we faced challenges, by staying determined and creative, we managed to overcome them. This experience gave us valuable lessons and made our experimental process stronger.

For this study we used the following Purification Protocol:

Bacterial Lysis: The first step involves breaking open the bacterial cells to release the proteins. This is done by resuspending the bacterial pellet in a lysis buffer(1× SGPP, 30 mM imidazole, 5% MOPS, 2% MgCl2), which helps to disrupt the cell membrane and degrade the cell wall. Following this, sonication is used to further break down the cell debris and ensure complete lysis.

1. **Capturing Purification**:

    - **Ni-NTA Column**: For 6-His tagged proteins, the lysate is first clarified by centrifugation, treated with benzonase nuclease, and then applied to a nickel-nitrilotriacetic acid (Ni-NTA) column. The Ni-NTA buffer consisted of a binding buffer (containing 50 mM sodium phosphate, 300 mM NaCl, and 10 mM imidazole) and an elution buffer (containing 50 mM sodium phosphate, 300 mM NaCl, and 250 mM imidazole). The peak fractions were collected, pooled and transferred to the next purification step (SEC)

    - **GST Column**: To purify a GST fusion protein, and at the same time to get rid of the GST-tag, first we bound the protein to washed and equilibrated Glutathione Sepharose column at 5°C. Then we washed the bound protein with 10 column volumes of Cleavage Buffer. Then, we mixed PreScission™ Protease (ThermoFisher Scientific®) with Cleavage Buffer and applied it to the protein-bound Sepharose column. Finally, we collected the eluate which contained the cleaved protein of interest while the GST tag and protease remained bound to the column. Peak elution fractions also were collected and transferred to SEC step.

2. **Polishing steps:** The final purification steps involved using size exclusion chromatography column (GE Superdex 75) with a HEPES-based buffer. The elution fractions were collected and analyzed; however, they were not pure enough for the downstream application, therefore we decided to add an extra step which included Q-column purification for Nsp-14 (Fig. 21-22) and S-column purification for Nsp-10(Data not shown). Protocols can be found at WWW.cytivascience.com.

Finally, after optimizing expression and purification process we isolated both Nsp 10 and Nsp14 proteins (Fig. 23).

Table 1. Contaminants in SEC Elution fractions after Nsp14 purification

Batch-Tag of Listed Accession Numbers

Rank	Uniq Pep	Acc #	06-10-2015_034/results1				Protein MW	Species	Protein Name
			Num Unique	% Cov	Best Disc Score	Best Expect Val			
1		K9N7C7	14	2.2	4.74	3.6e-8	789570.8	CVEMC	Replicase polyprotein 1ab
2		P0A6Y8	10	22.1	4.98	1.3e-8	69115.5	ECOLI	Chaperone protein DnaK
2-1	1	C5B7L7	2	4.6	3.05	3.0e-5	68509.8	EDWI9	Chaperone protein DnaK
3		P17169	11	26.1	4.10	5.6e-7	66894.9	ECOLI	Glutamine--fructose-6-phosphate aminotransferase [isomerizing]
4		P0A440	6	32.1	4.71	4.1e-8	31921.0	ECOL6	Formyltetrahydrofolate deformylase
5		P00760	6	28.0	4.46	1.2e-7	25785.5	BOVIN	Cationic trypsin
6		P0A894	4	22.2	4.95	1.5e-8	32492.6	ECOLI	UPF0042 nucleotide-binding protein YhbJ
7		P25888	4	10.8	3.13	1.3e-5	49989.7	ECOLI	ATP-dependent RNA helicase RhlE
8		Q12306	3	24.8	3.18	2.8e-5	11597.1	YEAST	Ubiquitin-like protein SMT3
9		P0A9K9	2	11.7	3.71	2.9e-6	20853.0	ECOLI	FKBP-type peptidyl-prolyl cis-trans isomerase SlyD
10		P00766	2	9.4	2.83	6.0e-5	25666.4	BOVIN	Chymotrypsinogen A
11		Q0AIY1	1	3.1	3.05	3.0e-5	69873.6	NITEC	Chaperone protein DnaK
12		Q52325	1	1.7	2.61	7.2e-5	73588.6	SALTY	Bifunctional polymyxin resistance protein ArnA
13		P04264	1	1.9	2.48	5.5e-4	66039.2	HUMAN	Keratin, type II cytoskeletal 1
14		P00763	1	4.1	2.27	5.8e-4	26228.0	RAT	Anionic trypsin-2
15		P64588	2	8.7	1.05	0.0021	23401.5	ECOLI	Transcriptional regulator YqjI
16		P0A951	2	12.9	1.16	3.5e-4	21887.2	ECOLI	Spermidine N(1)-acetyltransferase

Search Compare in ProteinProspector 5.14.1

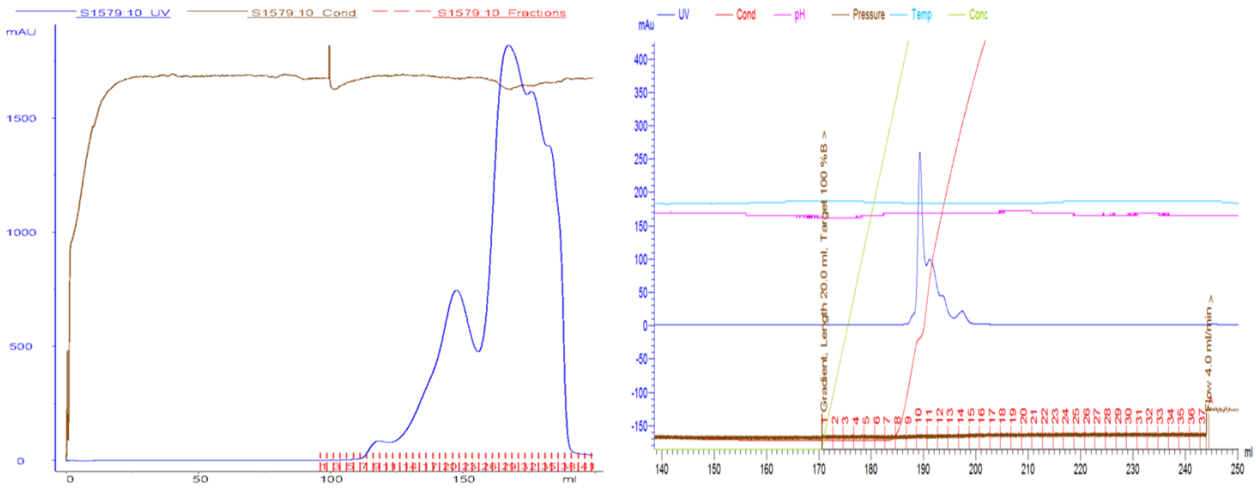


Figure 21. Left panel: SEC Profile of Nsp14, fractions 26-34 collected, elution vol 167 mL. Right panel: Q-column purification of pooled fractions 26-34, fractions 10,11,12 collected (see next figure)

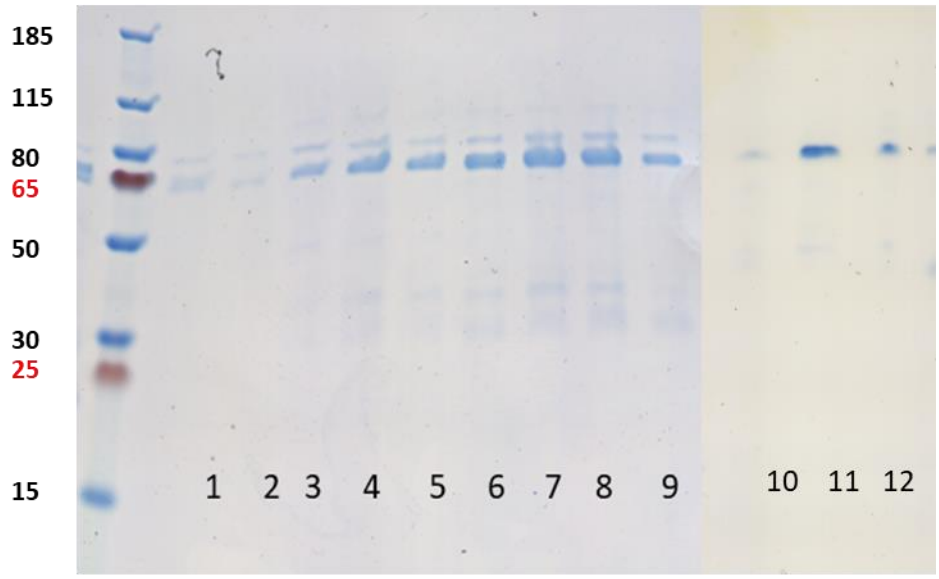


Figure 22. SDS PAGE GEL of fractions 26-34 of peak 2 (lane 1-9), lanes 10-12 represented peak fractions pooled and purified with Q column.

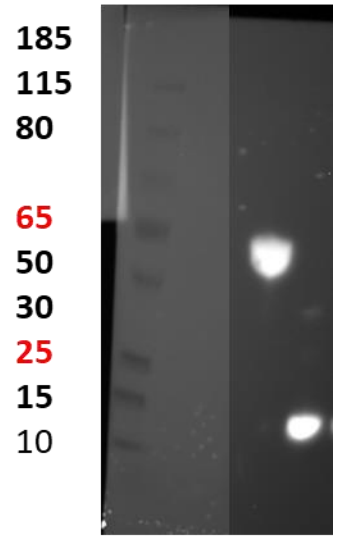


Figure 23. Western blot of Nsp 14 (below 65 KDa referred to MWL) and Nsp 10(Below 15 KDa referred to MWL), purified. Theoretical MW for Nsp 14 = 59KDa and MW for Nsp10 = 15.8 KDa

Example 3 Expression, purification and structural analysis of ADP-ribosylation factor, ARF1, from Entamoeba histolytica bound to Mg^{2+}–GDP

Entamoeba histolytica, the causative agent of amebiasis, a diarrheal illness leading to amoebic liver abscesses and amoebic colitis, affects around 50 million people worldwide. Despite only 10% of those infected developing symptomatic amebiasis, an estimated 100,000 deaths occur annually. The emergence of resistant parasite strains necessitates the search for effective treatments. ADP-ribosylation factor (ARF), a member of the GTP-binding protein family, is crucial in eukaryotic cells, regulating vesicular traffic and intracellular signaling by associating with cell membranes. The crystal structure of *E. histolytica's* ARF1, bound to magnesium and GDP at 1.8 Å resolution, has been determined. Comparative analysis with other eukaryotic ARF proteins reveals a highly conserved structure, supporting the inter-switch toggle mechanism for communicating conformational states to partner proteins (Serbzhinskiy et al., 2015).

The cloning, expression, and purification processes adhered to standard protocols from the Seattle Structural Genomics Center for Infectious Disease (SSGCID).

Cloning: The 174-residue gene of *E. histolytica* for putative ADP-ribosylation factor 1 (ARF1) was amplified from genomic DNA and inserted into an expression vector (pAVA0421), incorporating an N-terminal six-histidine affinity tag and a human rhinovirus 3C protease cleavage sequence using ligation-independent cloning (LIC). Gel-extracted and column-purified PCR inserts underwent treatment with T4 DNA polymerase in 96-well plates. These LIC-ready inserts were combined with LIC-ready AVA0421 vector and transformed into Escherichia coli hosts, followed by plating onto LB-Agar grills with antibiotic markers. Single colonies from each transformation were cultured overnight in LB with antibiotics in 96-well plates. Plasmid DNA was then isolated from the overnight cultures using 96-well plate plasmid mini-prep kits.

Expression and purification Protein expression follower the previously developed and published protocol (Bryan et al., 2011; Choi et al., 2011). Purification involved a four-step procedure, commencing with Ni2+-affinity chromatography (IMAC), followed by N-terminal His tag cleavage using 3C protease, subsequent removal of the cleaved tag via a second IMAC

step, and concluding with size-exclusion chromatography (SEC). All chromatography operations were executed on an ÄKTApurifier 10 (GE), utilizing automated IMAC and SEC protocols consistent with established methodologies (Bryan et al., 2011).

Thawed bacterial pellets were immersed in a beaker containing 200 ml of lysis buffer (1× SGPP, 30 mM imidazole, 5% MOPS, 2% MgCl2) and subjected to 15 minutes of sonication, alternating between 5-second pulses and 10-second pauses. Following sonication, the suspension was treated with Benzonase nuclease (Millipore) for 40 minutes at room temperature, then centrifuged at 10,000 rev min−1 for 1 hour using a Sorvall centrifuge (Thermo Fischer Scientific®). The resulting lysate, after centrifugation and filtration, underwent automated IMAC purification on an ÄKTApurifier 10 (Cytiva, former GE) equipped with a HisTrap HP 5 × 5 ml column (Cytiva, former GE). Peak fractions from the IMAC purification were collected and concentrated using an Amicon purification system (Millipore®).

N-terminal His tag cleavage was achieved through dialysis of the target protein with His-MBP-3C protease at 4°C overnight in 3C reaction buffer. The reaction mixture was then passed over a second Ni2+-affinity column to eliminate His-MBP-3C protease, uncleaved protein, and cleaved His tag. The cleaved protein, recovered in the flowthrough, was concentrated before being loaded onto the final SEC column (Superdex 75, Cytiva, former GE). Post-automated SEC run, peak fractions were gathered and subjected to SDS–PAGE analysis to confirm the presence of the desired protein. Comparison with molecular-weight standards indicated a decreased retention time on the SEC column, suggesting protein multimerization. The eluted peak fractions, falling within the molecular-mass range of 120–210 Da, implied the existence of multimeric complexes, possibly tetramers or larger (data not shown). These pooled peak fractions were concentrated, flash-cooled in liquid nitrogen, and stored at −80°C until utilized for crystallization.

The crystallization, X-ray data collection, and processing of purified E. histolytica ARF1 (13.5 mg / ml) involved screening in 96-well sitting-drop plates against the Wizard I and II crystal screens (Rigaku Reagents). Equal volumes of protein solution (0.4 µl) and precipitant solution were set up at 289.15 K against a reservoir (80 µl) in sitting-drop vapor-diffusion

format. The final crystallization precipitant was Wizard I condition No. 42, comprising 15% ethanol and 0.1 M Tris pH 7.0 at 289.15 K. Cryoprotection of crystals was achieved using crystallant with 25% ethylene glycol, followed by flash-cooling using liquid nitrogen. Data collection occurred at 100°C on Advanced Light Source beamline 5.0.3, utilizing an ADSC Quantum 315 CCD detector with 1° oscillations and a crystal-to-detector distance of 220 cm. Data were processed using XDS/XSCALE (Kabsch, 2010). Phases were determined through molecular replacement employing Phaser from the CCP4 suite of programs, with PDB entry 1r8s as a search model (McCoy, 2007; Winn et al., 2011). Structural refinement involved multiple cycles of [phenix.refine] program (Afonine et al., 2012), followed by manual rebuilding with Coot (Emsley et al., 2010). Structural integrity was assessed using MolProbity (Chen et al., 2010). Data-reduction and refinement statistics are provided in Tables 2 and 3 and are deposited in the PDB (PDB entry 4ylg). Figures, overlays, and electrostatic surface potentials were generated using PyMOL (v.1.5.0.4; Schrödinger).

Results:

ARF proteins, vital for cellular function, exhibit high sequence and structural similarity across diverse species, including *E. histolytica* (EhARF1). The conservation of guanine nucleotide-binding residues and hydrophobic core underscores evolutionary pressure to maintain protein fold. EhARF1 shares 77–83% identity with other taxa like plants, yeasts, mammals, and protozoa, with the closest relative being *E. invadens* (97% identity). Human ARF1 shares 79% sequence identity (Fig. 22). Structural analysis reveals significant similarity between GDP-bound EhARF1 (PDB entry 4ylg) and human ARF1 (PDB entry 1hur), with an r.m.s.d. of 0.5 Å between Cα atoms, indicating nearly identical architecture and substrate-binding centers.

Figure 24. Multiple sequence alignment shows a high similarity of ARF1 among eukaryotic species

Reference sequences are from E.histolytica (XP_654041.1), E. invadens (XP_004257850.1), Homo sapiens (NP_001649.1), Arabidopsis thaliana (NP_182239.1) and Saccharomyces cerevisiae (NP_010089.1). Multiple sequence alignment was conducted using Clustal Omega (Sievers et al., 2011) and the figure was generated using BoxShade v.3.21

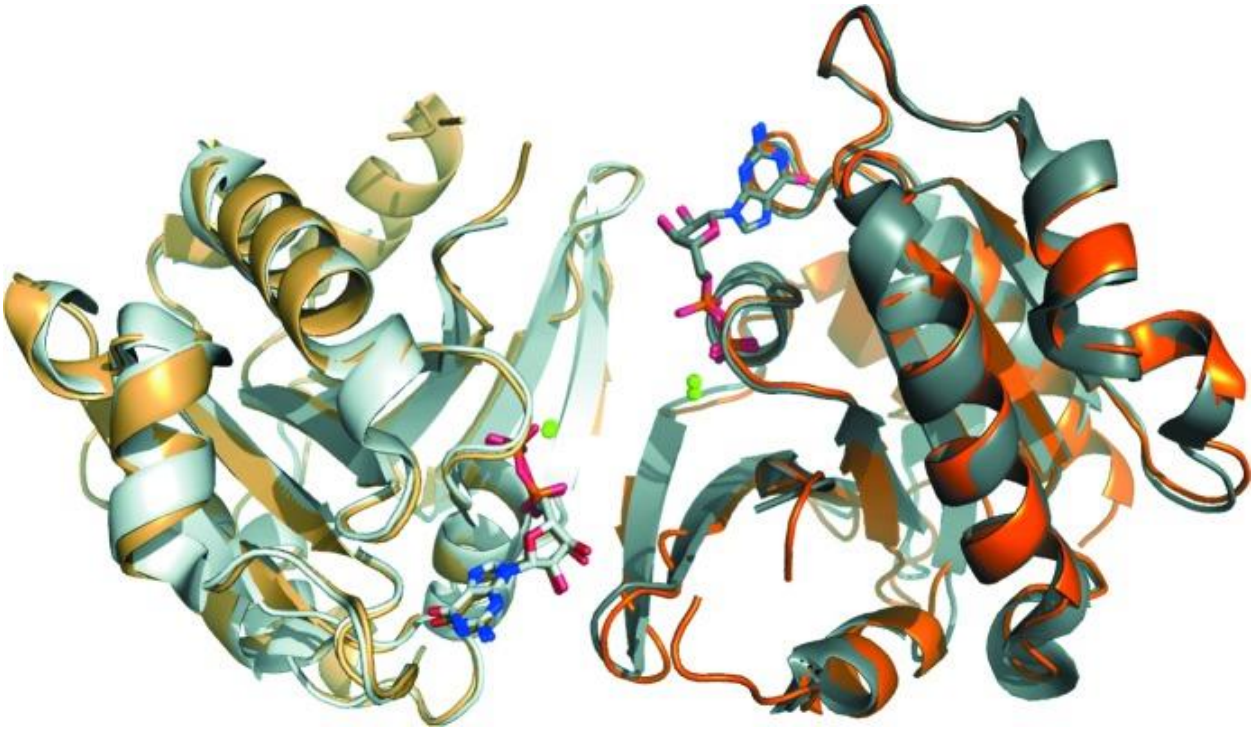


Figure 25. ARF1 from E. histolytica is structurally similar to human ARF1.

The structure of ARF1 from E. histolytica (gray) is overlaid with ARF1 from H. sapiens (PDB entry 1hur; orange) in the magnesium GDP-bound state (the superposition r.m.s.d. of PDB entries 4ylg and 1hur is 0.50 Å overall, calculated on all common Cα atoms). Similar to the human structure of ARF1, the interswitch region is largely disordered in the magnesium GDP-bound state.

A 1.8 Å resolution X-ray dataset was collected for the EhARF1–Mg2+–GDP structure, yielding a crystallographic R cryst of 22.3% and R free of 27.2%. The crystal's asymmetric unit in space group C2 comprises two molecules, oriented such that each molecule's GDP-binding site forms crystal-packing contacts via a β-sheet hydrogen-bonding pattern (Fig. 25). Each monomer features five α-helices, seven β-strands, and four 310-helices. Highly conserved amino acid residues, primarily located in the second α-helix and second/third β-strands, constitute the GDP-binding site (Bourne et al., 1991). The G-loops, crucial for small GTPases' functional activity, notably G1 and G4, exhibit higher conservation in ARF family proteins than secondary-structure units (Paduch et al., 2001). Key elements like Thr27 in the G1 loop play pivotal roles in Mg2+ ion binding. Switch I and II, within G2 and G3 loops respectively, contribute to protein conformational changes (Pasqualato et al., 2002). G4 and G5 loops

facilitate guanine nucleotide base binding, with specific residues involved in nucleotide interaction and base recognition (Sprang, 1997), Fig. 26.

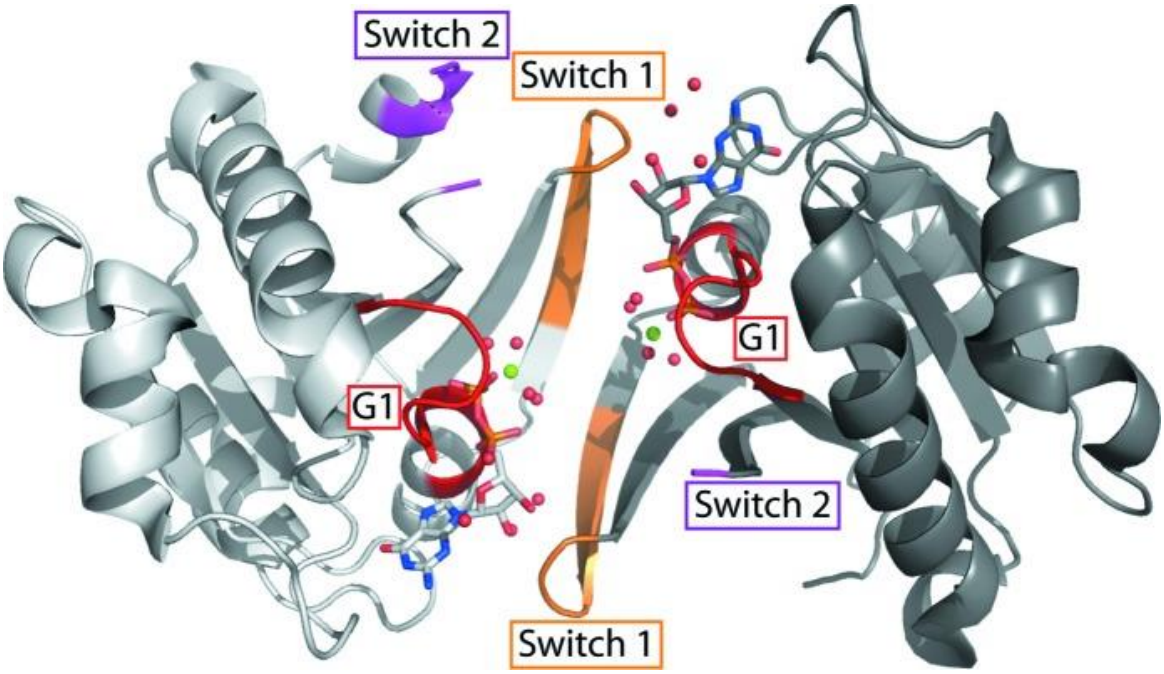


Figure 26. The structure of ARF1 from E. histolytica.

The protein forms a homodimer (light gray and dark gray). The G1 loop (also called the P--loop; red) creates a tight packing interaction surrounding the GDP and magnesium (green). Switch 1 (orange) creates the interface for the packing of the homodimer and the switch 2 region is only partially ordered (purple). There are multiple water-mediated contacts between ARF1, GDP and magnesium (water, red spheres; Mg2+, green spheres). Source: Serbzhinskiy et al., 2015

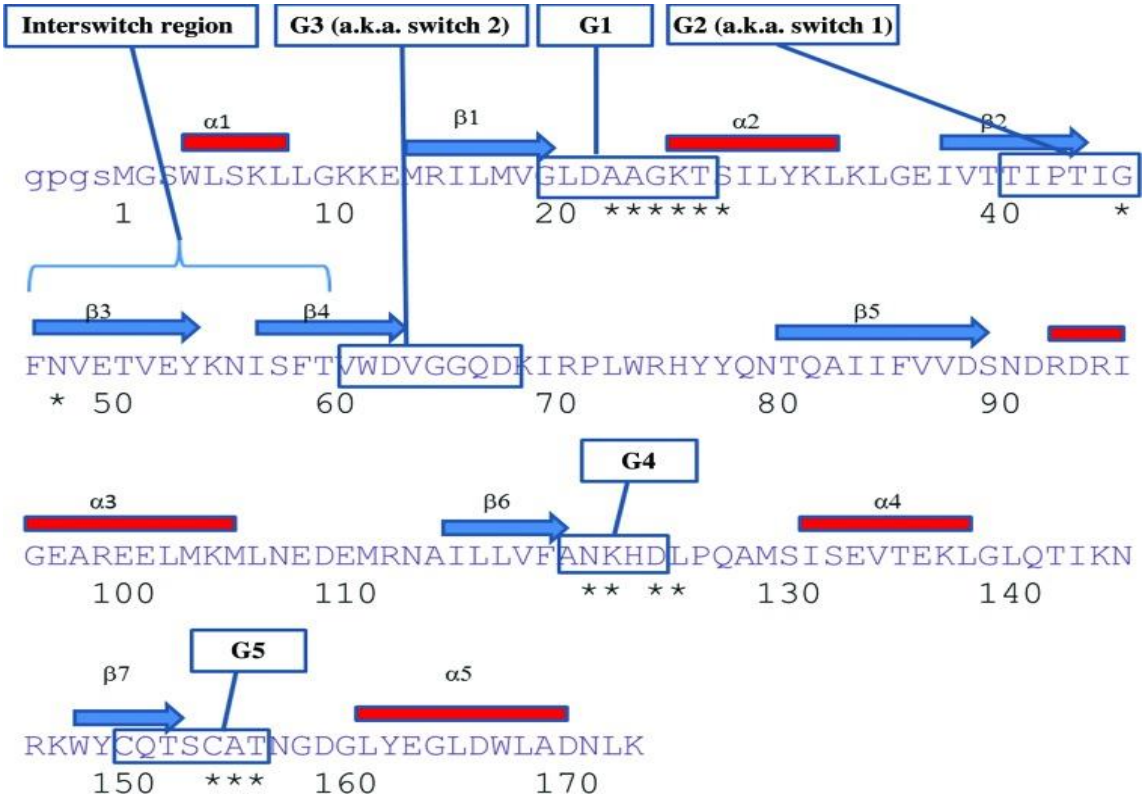


Figure 27. Structural elements of ARF1 from E. histolytica aligned with its primary structure.

Red boxes, α-helices; blue arrows, β-strands. G1–G5, G-loops. Residues directly responsible for GDP binding are marked with asterisks. Switches I and II and the inter-switch region are marked individually.

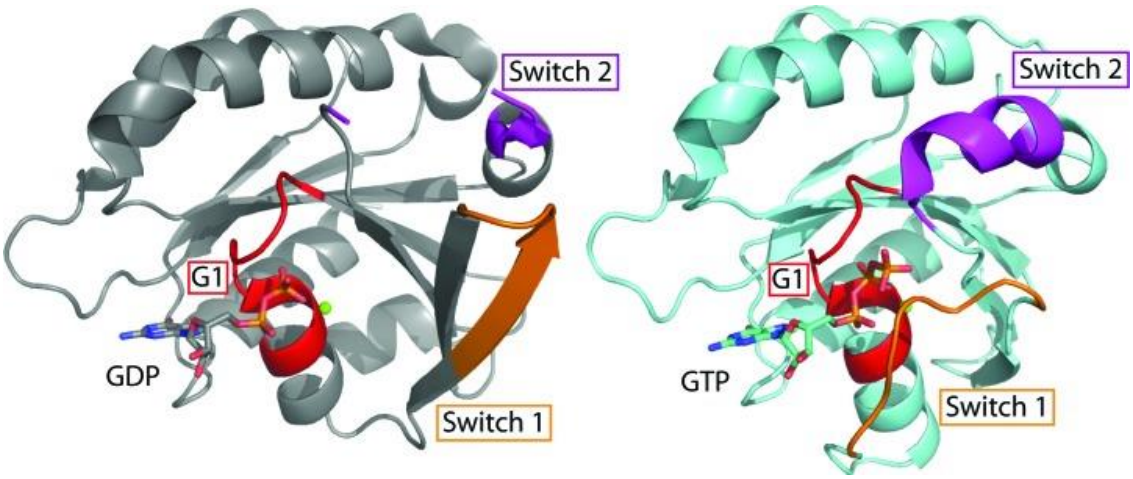


Figure 28. Comparison of ARF1 from E. histolytica bound to GDP (left, gray) with human ARF1 bound to GTP (PDB entry 4hmy; Ren et al., 2013 right, cyan).

The conformation of the P-loop (red) does not change upon GTP hydrolysis; however, there is a large conformational change in both the switch 1 (orange) and switch 2 (purple) regions. Although disordered in the GDP-bound structure, in the presence of GTP the switch 2 region adopts an α-helical conformation in the human ARF1 structure (right). Additionally, the switch 2 region that interacts with the γ-phosphate of GTP changes conformation to create a β-strand in the GDP-bound form.

Substrate recognition and comparison with GTP-bound human ARF1

ARF proteins, part of the small GTPases superfamily, exhibit two independent conformations when binding GTP and GDP. Their hydrolysis and exchange mechanisms rely on effector proteins (Vetter & Wittinghofer, 2001). Human ARF1–ARF6, active in their GTP-bound state, recruit coat proteins and form vesicles.

Comparing EhARF1–GDP with human ARF1's GTP-bound structure (PDB entry 1hur; Amor et al., 1994), we noted significant conformational differences (Fig. 28). In the GDP-bound form, EhARF1's switch I residues fold over, resembling an open mousetrap, while in human ARF1's GTP-bound state, they form a closed mousetrap. Such changes affect interaction with GTP/GDP-exchange proteins and GTPase-activating proteins (Donaldson & Jackson, 2000). Conformational shifts also impact the protein's inter-switch toggle mechanism (Pasqualato et al., 2002).

Example 4: Increasing structural coverage for tuberculosis drug targets

In the series of studies conducted in 2014 the group explored an intra-genus "homolog-rescue" approach aimed at enriching the structural data pool for TB drug discovery (Baugh et al., 2015). Out of 179 potential TB drug targets slated for x-ray structure analysis, merely 16 produced crystal structures. Incorporating 1675 homologs from nine different mycobacterial species into the investigation led to the resolution of structures representing 52 additional targets that were previously challenging to tackle. To assess the viability of these homolog structures in TB drug design, we conducted comparisons between the active sites of 106 pairs of *Mycobacteria tuberculosis* (Mtb) and non-TB mycobacterial enzyme homologs with experimentally determined structures, employing metrics of active site congruence, including the alignment of continuous pharmacophoric property distributions.

Comparing structures, 19 out of 22 pairs with >55% sequence identity showed highly similar active site features, suggesting preserved morphology and chemistry. Applying these results to 52 non-Tb mycobacterial structures, 41 shared >55% sequence identity with Mtb targets, tripling the structural coverage of 179 Mtb targets from 9% to 32%.

To support this work author participated in production 137 proteins from six *Mycobacterium* species. The complete list can be found in PDB.org >search field: “Serbzhinskiy”.

Case study 2.

Downstream process development including formulation for Monoclonal Antibodies (mAb) and vaccine product.

Example 1

Support study for stability of modified Plant Virus nanoparticles.

In a biotech company the author affiliated with, a successful attempt was made to create a conjugated COVID-19 vaccine. This vaccine consisted of a modified plant virion conjugated with the spike protein antigen. A modified plant virus nanoparticle with an N-terminal lysine mutation was produced by infecting wild-type *Nicotiana* plants with modified virions. Infected tissue was harvested, soluble protein was isolated, and virions were purified using a combination of Capto-Q and Capto-Core 700 chromatography. The modified virion was inactivated and sterilized by exposing the virus to ultraviolet light within an ISO 5 environment. Inactivation was confirmed through virus infectivity assays.

Subsequently, the plant virion was chemically conjugated to CoV-Receptor Binding Domain. However, during the production of the virion particles, they exhibited different properties compared to the wild type. One particulate issue was particles aggregation and disintegration (Fig. 29). These issues made the conjugation process more difficult and less productive, with up to 75% loss in the process. To address this problem, a new formulation development study was proposed.

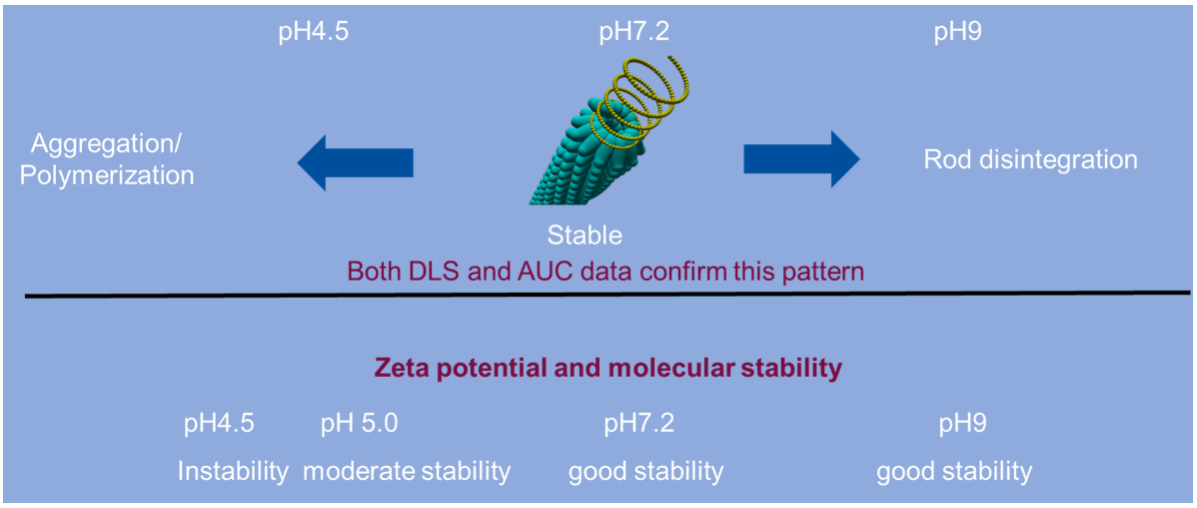


Figure 29. Effects of pH on the plant virion stability

Summary of Root Cause Analysis

The current conjugation process employs a PBS-based buffer system with a higher salt concentration. However, this buffer system may not be the most suitable for the current conjugation process. To address this, several key adjustments need to be made.

Firstly, we need to determine the optimal pH for the conjugation process (Fig. 30). The pH of the buffer system can significantly impact the efficiency and stability of the conjugation, so identifying the precise pH level that maximizes these factors is crucial.

Secondly, the ionic strength of the conjugation buffer needs to be optimized (Fig. 30). The current high salt concentration may be contributing to issues in the process, such as protein aggregation or reduced binding efficiency. By fine-tuning the ionic strength, we can enhance the stability and effectiveness of the conjugation reaction.

Lastly, an additional stabilizing component needs to be introduced into the solution (Fig. 30). This stabilizing agent can help prevent protein aggregation and maintain the integrity of the conjugated product. It is essential to identify a suitable stabilizing component that works well with the modified virion and the spike protein antigen.

In summary, to improve the conjugation process, we need to: determine the optimal pH, optimize the ionic strength of the buffer, and add an appropriate stabilizing component. These steps are critical to enhancing the efficiency and productivity of the conjugation process, ultimately leading to a more effective COVID-19 vaccine.

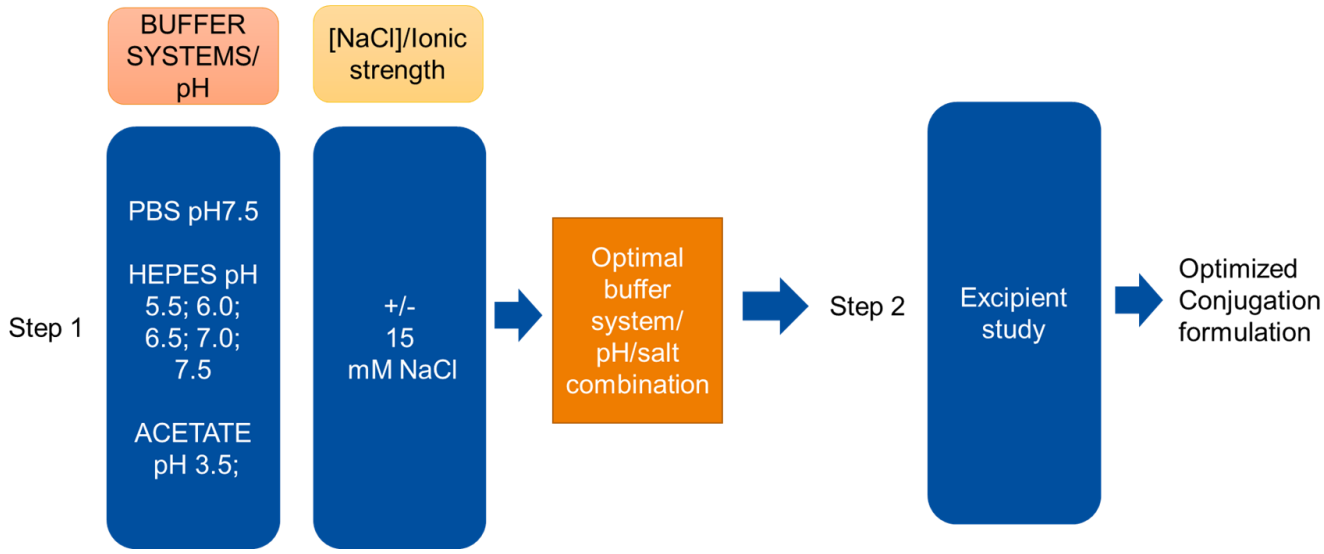


Figure 30. Screening study design

Testing plan

We tested three different buffer systems: Acetate, HEPES, and PBS, across a pH range of 3.5 to 7.5. These tests will be conducted with both no salt and low salt concentrations.

- Test all samples at T0 along with the virion standard.
- Conduct stress testing at +55°C after 24 hours.
- Evaluate all formulations after 2 weeks at +2 to +8°C.

Analytics for the Study:

- Particle size analysis using Differential Light Scattering (DLS).
- Measurement of compound Melting Temperature (Tm).
-
- Intrinsic Fluorescence maximum peak measurement.
- Evaluation of the solution's appearance to check for opalescence or cloudiness in an otherwise clear solution.

Results:

The results of the study will be presented in Table 2, and Fig. 31-32. Formulation stability under stress testing conditions will be primarily determined by analyzing particle size, appearance, and Intrinsic Fluorescence (IF) parameters.

At T0, DLS, DSF, and IF measurements are very similar, if not identical, except for the DLS results of the Acetate formulation. Notably, the Acetate solution was turbid immediately, suggesting that a pH of 3.5 is too acidic for the viral particles, causing rapid aggregation under these conditions.

After holding compound in 55C for 24 hours both PBS buffer systems and HEPES buffer without salt showed increased colloidal particle size by DLS and appearance (Table 2., Fig 31).

Table 2. Summary of formulation screening results.

	Sample Name	Concentration (mg/mL)			DLS (nm)			DSF - Tm(°C)			IF (Emission Max)			Appearance		
		Perform and Report			Perform and Report			Perform and Report			Perform and Report			Perform and Report		
		T 0	T 24hr@55 C	T 2wk @4C	T 0	T 24hr@55 C	T 2wk @4C	T 0	T 24hr@55 C	T 2wk @4C	T 0	T 24hr@55 C	T 2wk @4C	T 0	T 24hr@55 C	T 2wk @4C
1	PBS + 137mM NaCl pH 7.5	3.45	5.22	3.43	59.8	330	59.9	75.1	77.2	75.2	325nm	325nm	325nm	Clear	Turbid	Clear
2	PBS + 15mM NaCl pH 7.5	3.55	6.05	3.55	43.9	1342.2	42.1	74.9	76.8	75.4	325nm	330nm	325nm	Clear	Turbid	Clear
3	HEPES pH 7.5	4.01	4.3	4.02	44.6	54.4	43	75.3	75.7	75.9	325nm	325nm	325nm	Clear	Slight Turbidity	Clear
4	HEPES + 15mM NaCl pH 7.5	4.01	4.47	4.02	44.8	68.6	42.1	75.4	75.8	75.4	325nm	325nm	325nm	Clear	Clear	Clear
5	HEPES pH 7.0	3.75	3.78	3.77	44.7	42.2	43.3	75.4	75.4	75.7	325nm	325nm	325nm	Clear	Clear	Clear
6	HEPES + 15mM NaCl pH 7.0	3.74	3.8	3.74	44.3	45.2	43	75.4	75.3	76	325nm	325nm	325nm	Clear	Clear	Clear
7	HEPES pH 6.5	3.26	3.31	3.27	44.4	40.9	43.9	75.4	75.4	75.7	325nm	325nm	325nm	Clear	Clear	Clear
8	HEPES + 15mM NaCl pH 6.5	3.29	3.3	3.25	44.6	44.1	43.3	75.4	75.2	75.7	325nm	325nm	325nm	Clear	Clear	Clear
9	HEPES pH 6.0	5.18	5.25	5.34	42.7	42.6	42.5	75.4	75.4	75.7	325nm	325nm	325nm	Clear	Clear	Clear
10	HEPES + 15mM NaCl pH 6.0	5.26	5.25	5.29	39.8	42.1	40.7	75.5	75.3	75.8	325nm	325nm	325nm	Clear	Clear	Clear
11	HEPES pH 5.5	4.42	4.48	4.41	49.2	127.8	46.5	75.3	75.7	75.7	325nm	325nm	325nm	Clear	Clear	Clear
12	HEPES + 15mM NaCl pH 5.5	4.38	4.74	4.39	45.4	85.6	44.8	75.4	75.7	75.8	325nm	325nm	325nm	Clear	Clear	Clear
13	Acetate pH 3.5	6.65			261.2			77.3			320nm			Turbid		
14	Acetate + 15mM NaCl pH 3.5	7.13			263.1			77.1			325nm			Turbid		
NA	20TMV003 Control - No UV	0.57	0.55	0.45	44.8			77.9		77.9	325nm		325nm	Clear		Clear

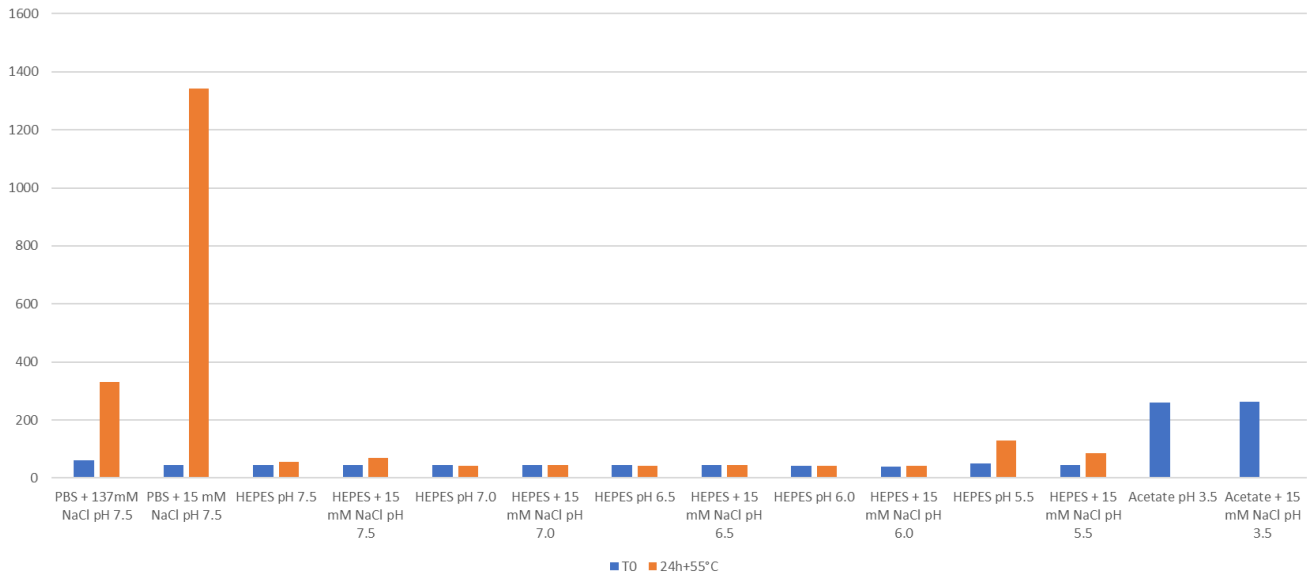


Figure 31. Particle size analysis, by DLS (nm)

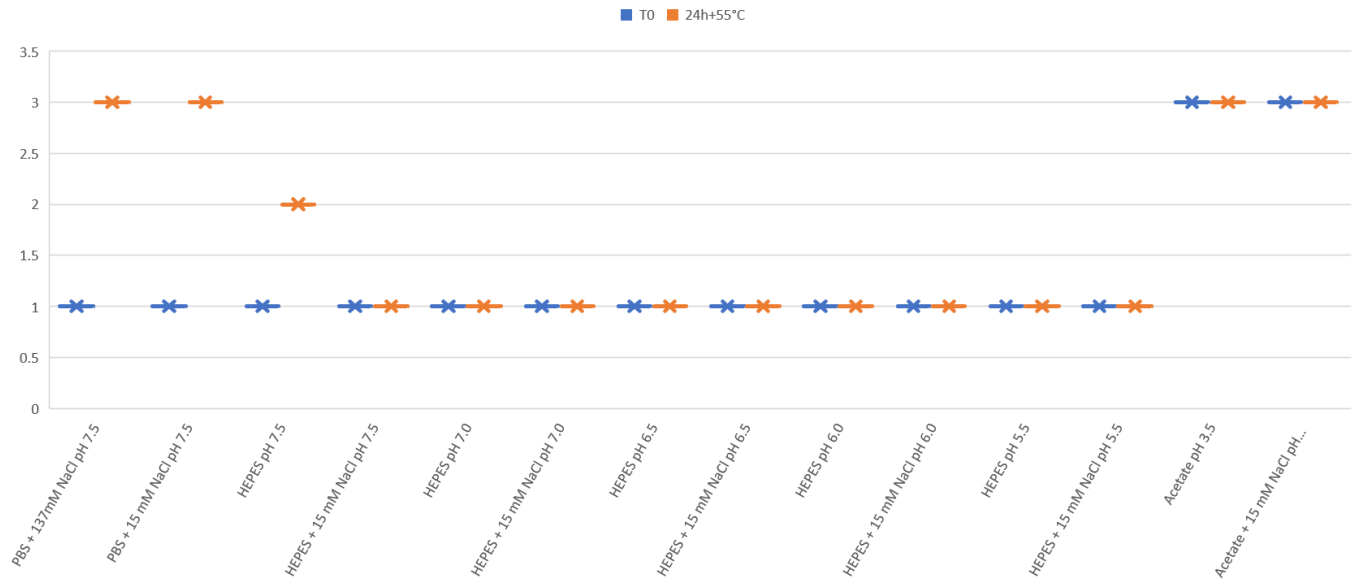


Figure 32. Virion Formulation Appearance (1- clear, 2 – slight turbidity, 3- turbid)

Stress testing has highlighted critical points related to potential protein aggregation, particularly when exposed to higher salt concentrations and at pH levels below 6.0 and above 7.0. These conditions can significantly affect the stability of the formulation,

necessitating a comprehensive analysis to mitigate aggregation risks. Addressing this issue is crucial for ensuring the efficacy and safety of the final product.

To streamline the production process and ensure the final formulation's compatibility with intravenous (IV) injection systems, testing an additional buffer system, specifically histidine (His), was completed and showed a promising result. The histidine buffer system offered a more stable environment, potentially reducing the risk of aggregation and improving overall formulation stability under various conditions.

Moreover, a detailed excipient study is essential to further enhance the stability of the formulation. Excipients play a vital role in maintaining the structural integrity of the active pharmaceutical ingredient (API) by preventing degradation and aggregation. By carefully selecting and evaluating excipients, the formulation can achieve improved stability and a longer shelf life.

In addition to these studies, it is necessary to explore additional temperature parameters and conduct an extensive timepoint study. Understanding how the formulation behaves under different temperature conditions over time will provide valuable insights into its stability profile. This information is crucial for determining optimal storage conditions and ensuring the formulation remains stable throughout its intended shelf life.

Finally, incorporating knowledge gained from the initial screening studies into the conjugation process development is essential. These preliminary studies provide a foundation of understanding that can guide the optimization of the conjugation process. By leveraging this knowledge, the development team can refine the process to enhance efficiency, yield, and product quality.

Overall, a multifaceted approach involving stress testing, buffer system evaluation, excipient studies, temperature and timepoint analysis, and knowledge integration from screening studies is critical for developing a robust and stable formulation suitable for IV injection systems.

Example 2

Formulation Development of Antiviral Monoclonal Antibody (Mab)

Another experiment aimed to develop a formulation for an antiviral antibody that ensures the stability and efficacy of this API (Active Pharmaceutical Ingredient). The development plan is presented in Fig. 33.

For prescreening, 32 samples were prepared: 8 samples to test different pH levels and salt concentrations, 8 to test the addition of different amino acids, 8 to test the addition of sucrose, and 8 to test whether adding hydroxypropyl-beta-cyclodextrin (HPBCD) improves formulation stability. The analytical plan included prescreening tests, selecting the most stable formulations, and testing those formulations under stress conditions (Fig. 34).

Among the various proposed test methods, we chose to measure the melting temperature (Tm) of the solutions, use ELISA for activity testing, and perform particle size and polydispersity analysis by Dynamic Light Scattering (DLS). Polydispersity, characterized by the Polydispersity Index (PDI), indicates the uniformity of particle sizes within a sample, with lower values signifying more uniform sizes and higher values indicating a broader range of sizes. Particle size and PDI can reveal early aggregation patterns, while Tm can indicate conformational stability.

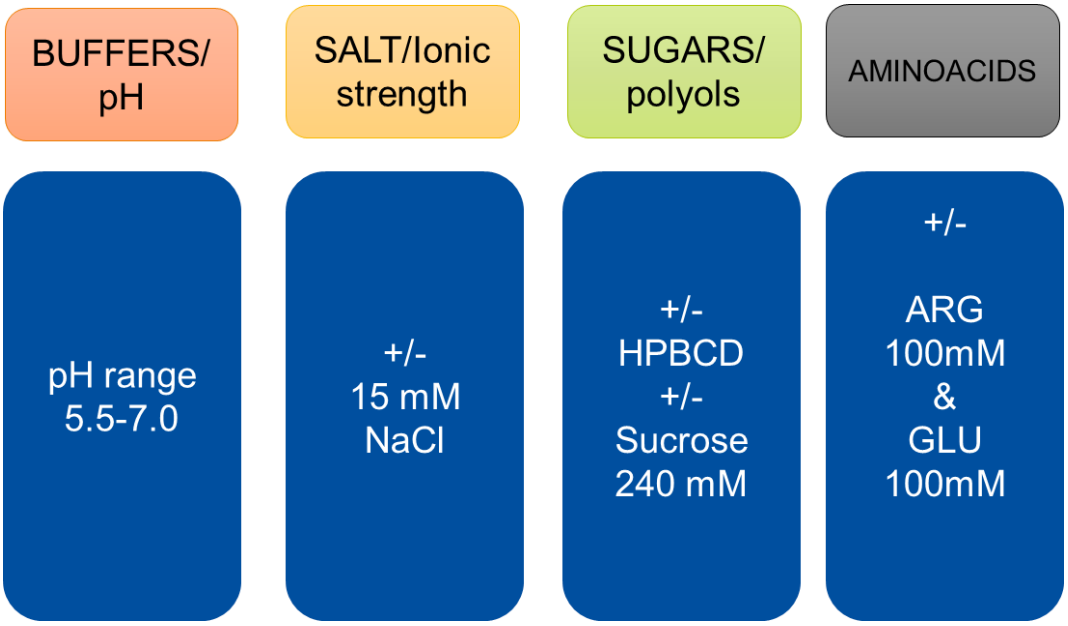


Figure 33. Workflow for formulation prescreening

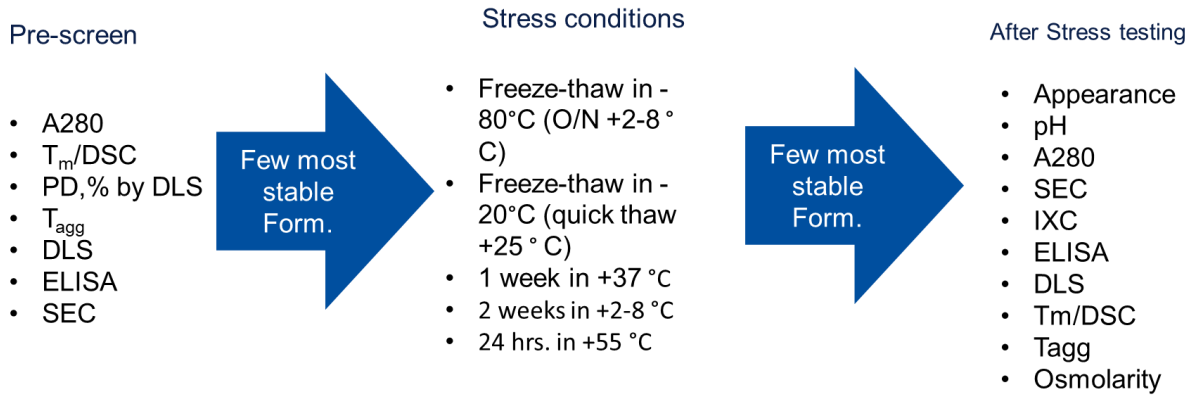


Figure 34. Analytical plan proposal for formulation development of Antiviral MAb (HSV8)

Study results

Analyzing Tm and polydispersity together, as presented in Fig. 35, is a helpful tool for selecting more stable formulations. The table shows the effects of salt concentration and the addition of multiple excipients while analyzing formulations across a wide pH range (5.5-

7.0). Adding salt can reduce polydispersity (tendency to aggregate) and adding amino acids and sugars can also contribute to the formulation's stability.

Among these results, sucrose showed good potential for maintaining conformational stability (measured by Tm), so formulations with sucrose were screened for functional activity and tendency to aggregate (Table 3). Unfortunately, all of them exhibited either higher PDI or lower ELISA numbers (indicated in red font), which points to a higher tendency to aggregate or inactivity (for this mAb, ELISA EC50 > 3 is considered inactive).

Therefore, for stress testing, we selected the best-performing formulations from the initial screening and subjected them to stress conditions of 55°C for 2 days and 37°C for 1 week. These formulations had pH levels of 6.5 and 7.0 with no salt. We tested these samples without any excipients and with combinations of them at 55°C for 24 hours and 37°C for a week (Table 4, 5). After this testing, the results were clear: the addition of HPBCD to both formulations showed the best overall stability, with the lowest PDI, smallest DLS radius, and highest ELISA readings.

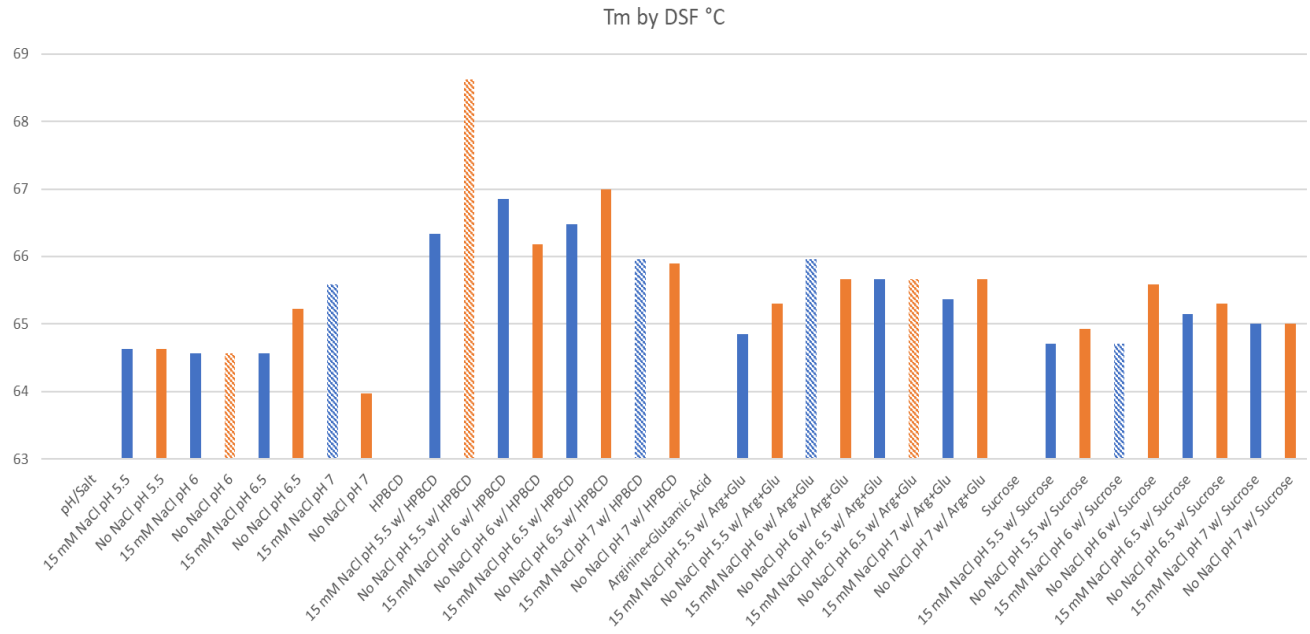


Figure 35. Tm by DSF, blue(+pattern) - 15mM NaCl added, orange + pattern - no salt. Pre-screening. Textured bars represent samples with the lowest polydispersity.

Table 3. Sucrose Formulation, prescreening

Sample Description	Conc, mg/mL	Tm by DSF °C	DLS Radius (nm)	DLS PD%	ELISA EC50	pH
15 mM NaCl pH 5.5 w/ Sucrose	17.36	64.70/89.75	5.3	33	3.737	5.82
No NaCl pH 5.5 w/ Sucrose	17.87	64.93/89.83	5.2	49.6	4.23	5.77
15 mM NaCl pH 6 w/ Sucrose	17.38	64.70/89.75	5.1	23.2	3.668	6.15
No NaCl pH 6 w/ Sucrose	17.67	65.59/89.83	7.3	56.6	1.615	6.07
15 mM NaCl pH 6.5 w/ Sucrose	18.16	65.15/89.75	6.6	46.3	2.392	6.58
No NaCl pH 6.5 w/ Sucrose	18.71	65.30/89.83	5.1	54.5	1.876	6.53
15 mM NaCl pH 7 w/ Sucrose	17.16	65.00/89.75	6.5	56.3	1.915	6.65
No NaCl pH 7 w/ Sucrose	17.51	65.00/89.68	5.2	49.6	2.991	6.56

Table 4. Stress testing, excipient combinations, 48 hours holding at +55C

Sample Description	Conc, mg/mL	Tm by DSF °C	DLS Radius (nm)	DLS PD%	ELISA EC50	pH
Condition 1 (pH 7.0) No Excipients, T=2 days, 55°C 24h	20.59	66.83/89.46	110.6	57.1	2.855	6.81
Condition 2 (pH 6.5) No Excipients , T=2 days, 55°C 24h	20.69	66.58/89.56	9.8	57.1	3.463	6.36
Condition 1 (pH 7.0) w/ HPBCD+Arg/Glu, T=2 days, 55°C 24h	20.63	66.39/89.76	11.4	57.1	1.287	6.94
Condition 2 (pH 6.5) w/ HPBCD+Arg/Glu, T=2 days, 55°C 24h	20.57	66.48/89.56	101.3	57.1	1.275	6.36
Condition 1 (pH 7.0) w/ HPBCD, T=2 days, 55°C 24h	20.72	66.39/89.56	105.1	57.1	2.549	6.79
Condition 2 (pH 6.5) w/ HPBCD, T=2 days, 55°C 24h	20.76	66.73/89.66	11.7	57.1	2.482	6.37

Table 5. Stress testing, excipient combination 1 week holding at +37C

Sample Description	Conc, mg/mL	Tm by DSF °C	DLS Radius (nm)	DLS PD%	ELISA EC50	pH
Condition 2 (pH 6.5) No Excipients , T=1 Week, 37°C	21.2	66.69/89.52	33	57.1	2.548	6.76
Condition 1 (pH 7.0) w/ HPBCD+Arg/Glu, T=1 Week, 37°C	35.31	67.08/89.66	266.7	57.1	1.507	7.32
Condition 2 (pH 6.5) w/ HPBCD+Arg/Glu,T=1 Week, 37°C	25.27	66.49/89.57	414.8	57.1	1.415	7.12
Condition 1 (pH 7.0) w/ HPBCD, T=1 Week, 37°C	21.03	66.84/89.37	10.9	57.1	2.272	7.02
Condition 2 (pH 6.5) w/ HPBCD, T=1 Week, 37°C	20.69	66.10/89.42	17.2	57.1	1.817	6.58

Summary

In this study, we aimed to develop a stable and efficacious formulation for an antiviral antibody. By analyzing Tm and polydispersity, we identified key factors influencing formulation stability. Salt and various excipients, including amino acids and sugars, were tested for their effects on stability across a pH range of 5.5-7.0. While sucrose showed potential for maintaining conformational stability, it ultimately resulted in higher aggregation or reduced activity.

The stress testing of selected formulations revealed that HPBCD addition significantly improved stability. Formulations at pH 6.5 and 7.0, tested under stress conditions, demonstrated that HPBCD provided the best overall stability, as indicated by the lowest PDI, smallest DLS radius, and highest ELISA readings. These findings highlight the importance of excipient selection in developing stable antibody formulations and provide a clear direction for future formulation development.

Discussion

The main goal of this dissertation is to outline a conceptual framework for developing high-throughput protein production pipelines. This involves addressing two primary challenges: selecting the target protein and designing the production pipeline. The selection of the target protein is crucial and depends on the researcher's or application scientist's objectives, whether it is a potential drug target, activity inhibitor, antibody, or a product for cosmetic or food industries.

The significance of the product is a major consideration. For instance, can the target protein help overcome drug resistance or treat rare diseases? This consideration influences the selection process, guiding researchers toward proteins with significant therapeutic or commercial potential.

Another critical perspective is the Chemistry, Manufacturing, and Controls (CMC) viewpoint. This involves evaluating whether the target protein is theoretically easy to produce and identifying potential challenges in the production line. Issues such as poor expression levels, formation of inclusion bodies, difficulties in purification due to physicochemical properties, or degradation post-purification must be considered. Additionally, the market potential and economic viability of producing the protein are essential factors in the decision-making process.

In this dissertation, the author addresses several of these questions, providing insights into the complexities of high-throughput protein production and offering strategies to overcome these challenges. By considering both the scientific and practical aspects of protein production, this framework aims to facilitate the efficient and effective development of protein production pipelines.

Below we will discuss key findings for each example given in the results section of this work and summarize their significance for establishing a robust and reproducible protein production pipeline.

Rho9 project:

To address cloning errors such as frame-shift mutations, we redesigned and re-done cloning methods. After that, all six clones were successfully amplified, purified, and cloned into the Rosetta *E. coli* strain. However, expression testing indicated that the protein band in the soluble fraction was significantly smaller than in the whole cell lysate, suggesting solubility issues. By optimizing the temperature during induction, we improved the solubility and functionality of the expressed protein. Western blot analyses demonstrated that these optimized conditions successfully produced functional, soluble proteins from all clones using both autoinduction media and IPTG induction methods. This structured study plan effectively addressed cloning errors and optimized protein expression, enhancing our understanding of the target protein and its potential applications.

Therefore, strategy of isolating active center of the protein clone and purify it separately was successful, especially for structure solve purposes.

SARS-CoV project

In this study, we aimed to express and purify two key RNA-synthesizing proteins, nsp10 and nsp14 from SARS-CoV, in Escherichia coli. We genetically modified nsp10 with an N-terminal GST tag and nsp14 with a C-terminal His tag and cloned these constructs into the pGEX-6p-1 and pRSFDuet-1 vectors, respectively. Transformation into Rosetta strain E. coli cells followed by culturing on antibiotic-containing plates confirmed successful transformation. Consistent expression and solubility across clones indicated reliable and reproducible protein production.

Despite initial success, we faced challenges such as low protein expression with the BL21 strain, which was resolved by switching to the LOBSTR strain. Unexpected DnaK chaperone contamination, identified via mass spectrometry, was eliminated by using the BL21(DE3) ΔdnaK strain. Our purification protocol involved bacterial lysis, affinity chromatography (Ni-NTA for nsp14 and GST column for nsp10), and size exclusion chromatography. Additional Q-column and S-column purification steps significantly improved protein purity.

Through optimization of expression and purification processes, we successfully isolated high-quality, soluble nsp10 and nsp14 proteins. Overcoming challenges such as low expression and contamination underscored the importance of flexibility and meticulous protocol adjustments. This work not only enhances our understanding of these proteins but also establishes a robust, reproducible pipeline for future SARS-CoV protein studies.

Entamoeba histolytica project

This project, upon successful completion of cloning, expression, and purification, aims to shed light on the mechanism of action of ARF1 and its similarity to other ARF proteins within the small GTPases superfamily. These proteins exhibit two independent conformations when binding GTP and GDP. Their hydrolysis and exchange mechanisms depend on effector proteins. Human ARF1–ARF6, active in their GTP-bound state, recruit coat proteins and form vesicles.

Comparing EhARF1–GDP with the GTP-bound structure of human ARF1 (PDB entry 1hur; Amor et al., 1994), we observed significant conformational differences (Fig. 3). In the GDP-bound form, EhARF1's switch I residues fold over like an open mousetrap, while in the GTP-bound state of human ARF1, they form a closed mousetrap. These changes affect interactions with GTP/GDP-exchange proteins and GTPase-activating proteins (Donaldson & Jackson, 2000). Additionally, conformational shifts impact the protein's inter-switch toggle mechanism (Pasqualato et al., 2002).

Conjugated viral vaccine project:

This study highlights the critical factors affecting the stability of the formulation under stress testing conditions, with a particular focus on particle size, appearance, and Intrinsic Fluorescence (IF) parameters. Initial findings indicate that acidic conditions (pH 3.5) lead to rapid aggregation, while elevated temperatures (55°C) exacerbate particle size increases in certain buffer systems.

The evaluation of the histidine buffer system demonstrated improved stability, suggesting its potential as a preferred buffer for reducing aggregation risks. Detailed excipient studies are necessary to enhance stability further, ensuring the structural integrity of the API and extending the formulation's shelf life.

Temperature and timepoint analyses are crucial for determining optimal storage conditions, providing insights into the formulation's long-term stability. Integrating these findings into the conjugation process development will guide the optimization efforts, enhancing the efficiency, yield, and quality of the final product.

Overall, a comprehensive approach involving stress testing, buffer evaluation, excipient selection, and long-term stability studies is essential for developing a robust and stable formulation suitable for intravenous (IV) injection systems.

Antiviral Monoclonal Antibody formulation development

In this study, our goal was to develop a stable and effective formulation for an API. Through the initial screening, we identified key factors that influence formulation stability. Various salts and excipients, including amino acids and sugars, were tested across a whide pH to assess their effects on API stability. Some excipients showed promise in maintaining conformational stability, however, they ultimately led to increased aggregation or reduced activity.

Stress testing of selected formulations (after screening) revealed that the addition of hydroxypropyl beta cyclodextrin (HPBCD) significantly enhanced stability. Final Formulations was then determined, and when subjected to stress conditions, it demonstrated that superior overall stability, as evidenced by the lowest polydispersity index, the smallest dynamic light scattering radius, and the highest activity. These findings underscore the critical role of excipient selection in developing stable antibody formulations and offer a clear direction for future formulation development.

Summary

Development of the downstream protein production pipeline is a complex multicomponent task with multiple challenges researchers meet along the way.

One of the progressive approaches applied in academia and Biotech industry is platform approach. It refers to a standardized method or set of procedures that can be applied to purify different proteins efficiently. This approach leverages common technologies, equipment, and protocols to streamline the purification process, making it more reproducible and scalable.

Here we have two main challenges – target protein selection and pipeline design.

Before the beginning of the process defining of the target protein or proteins needs to be done, and it may not be a trivial task. Both manufacturing and user-end aspects need to be considered. I.e. questions like "why do we need this product"? or "how can we make sufficient amount of it?" should be asked. After target was chosen researchers need to assess the property of the product ought to be produced. For that significant amount of research and/or in-silico investigations need to be done. For example, existing structures of orthologs/paralogs needs to be assessed. Prediction of biophysical properties such as pI will help streamline the manufacturing process.

Target selection process in SSGCID (Myler et al., 2009) can be used as a good example for such approach. There, proteins were grouped by important biological roles for example:

- Proteins involved in pathogenesis.
- Proteins essential to the pathogen's life and reproductive cycle.
- Proteins involved in antimicrobial/drug resistance.

Then series of bioinformatic and manual filters were applied using positive and negative selection criteria, such as sequence similarity to known drug target and biophysical characteristics such as solubility, amino acid composition and presence of transmembrane domain.

After research and properties prediction we need to design production pipeline. It starts with **cloning and expression screening** when we choose correct expression host and optimize cloning process (Case Study1). Often, the host chosen for the whole platform works just fine, but sometime, for example with the case of nsp10-nsp14 process customized clones with some gene deficiency or additions are needed. Sometime additional genetic engineering steps may be needed such as codon optimization, promoter selection, or signal peptide addition. Transformation efficiency and expression conditions (such as temperature and induction methods) also should be optimized.

After desirable expression level and solubility of the product are reached we may need to design post expression optimization and purification methods. They include cell lysis conditions, cell harvest timing, decision about using certain affinity tags or move forward with different capturing purification method, potential polishing method and final buffer / conditions in which the product will be stored.

If the expression and/or solubility of the product is not sufficient to reach researcher's goal then different "rescue" strategies may be applied including but not limited changing the expression system, modification of expression vector, host strain optimization, refolding from inclusion bodies, etc. This so called "feedback loop" or decision-making mechanism can be applied to all stages of protein production.

Overall platform approach implies starting with previously developed broad method and if it is not working overcome it with optimization strategy.

Optimizing critical process parameters are very important during development protein production pipeline. Choosing correct vector, host, expression and purification methods together with correct analytics for each step of the process is critical for assuring the project success.

In the case study 2 we see illustration of platform approach to formulation development for vaccine and antibody production. As we specified earlier, we started with more broad method and optimize it while moving forward. Most useful development tools here are high

throughput screening, stepwise formulation development approach and design of experiments (DOE).

In conclusion, the development of a downstream protein production pipeline is a multifaceted and challenging endeavor, requiring meticulous planning and optimization. The platform approach, widely adopted in academia and the biotech industry, offers a standardized framework that enhances efficiency, reproducibility, and scalability. This approach proved to be a valuable instrument for producing and upscaling quality protein products.

List of figures

List of tables

References

Afonine, P. v., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., & Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D: Biological Crystallography*, *68*(4). https://doi.org/10.1107/S0907444912001308

Altmann, F., Staudacher, E., Wilson, I. B. H., & März, L. (1999). Insect cells as hosts for the expression of recombinant glycoproteins. In *Glycoconjugate Journal* (Vol. 16, Issue 2). https://doi.org/10.1023/A:1026488408951

Asada, H., Uemura, T., Yurugi-Kobayashi, T., Shiroishi, M., Shimamura, T., Tsujimoto, H., Ito, K., Sugawara, T., Nakane, T., Nomura, N., Murata, T., Haga, T., Iwata, S., & Kobayashi, T. (2011). Evaluation of the Pichia pastoris expression system for the production of GPCRs for structural analysis. *Microbial Cell Factories*, *10*. https://doi.org/10.1186/1475-2859-10-24

Aslanidis, C., & de Jong, P. J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Research*, *18*(20). https://doi.org/10.1093/nar/18.20.6069

Asojo, O. A., Dranow, D. M., Serbzhinskiy, D., Subramanian, S., Staker, B., Edwards, T. E., & Myler, P. J. (2018). Crystal structure of chorismate mutase from Burkholderia thailandensis. *Acta Crystallographica Section F: Structural Biology Communications*, *74*(5). https://doi.org/10.1107/S2053230X1800506X

Ayyar, V. (2011). History of growth hormone therapy. *Indian Journal of Endocrinology and Metabolism*, *15*(7). https://doi.org/10.4103/2230-8210.84852

Baeshen, M. N., Al-Hejin, A. M., Bora, R. S., Ahmed, M. M. M., Ramadan, H. A. I., Saini, K. S., Baeshen, N. A., & Redwan, E. M. (2015). Production of biopharmaceuticals in E. Coli: Current scenario and future perspectives. In *Journal of Microbiology and Biotechnology* (Vol. 25, Issue 7). https://doi.org/10.4014/jmb.1412.12079

Baghban, R., Farajnia, S., Rajabibazl, M., Ghasemi, Y., Mafi, A. A., Hoseinpoor, R., Rahbarnia, L., & Aria, M. (2019). Yeast Expression Systems: Overview and Recent Advances. In *Molecular Biotechnology* (Vol. 61, Issue 5). https://doi.org/10.1007/s12033-019-00164-8

Bally, J., Nakasugi, K., Jia, F., Jung, H., Ho, S. Y. W., Wong, M., Paul, C. M., Naim, F., Wood, C. C., Crowhurst, R. N., Hellens, R. P., Dale, J. L., & Waterhouse, P. M. (2015). The extremophile Nicotiana benthamiana has traded viral defence for early vigour. *Nature Plants*, *1*. https://doi.org/10.1038/nplants.2015.165

Barta, A., Sommergruber, K., Thompson, D., Hartmuth, K., Matzke, M. A., & Matzke, A. J. M. (1986). The expression of a nopaline synthase - human growth hormone chimaeric gene in transformed tobacco and sunflower callus tissue. *Plant Molecular Biology*, *6*(5). https://doi.org/10.1007/BF00034942

Baugh, L., Phan, I., Begley, D. W., Clifton, M. C., Armour, B., Dranow, D. M., Taylor, B. M., Muruthi, M. M., Abendroth, J., Fairman, J. W., Fox, D., Dieterich, S. H., Staker, B. L., Gardberg, A. S., Choi, R., Hewitt, S. N., Napuli, A. J., Myers, J., Barrett, L. K., … Myler, P. J. (2015). Increasing the structural coverage of tuberculosis drug targets. *Tuberculosis*, *95*(2). https://doi.org/10.1016/j.tube.2014.12.003

Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H., & Kompala, D. S. (1990). Plasmid-encoded protein: The principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnology and Bioengineering*, *35*(7). https://doi.org/10.1002/bit.260350704

Biddick, R., & Young, E. T. (2009). The disorderly study of ordered recruitment. In *Yeast* (Vol. 26, Issue 4). https://doi.org/10.1002/yea.1660

Botella, L., Vaubourgeix, J., Livny, J., & Schnappinger, D. (2017). Depleting Mycobacterium tuberculosis of the transcription termination factor Rho causes pervasive transcription and rapid death. *Nature Communications*, *8*. https://doi.org/10.1038/ncomms14731

Bourne, H. R., Sanders, D. A., & McCormick, F. (1991). The GTPase superfamily: conserved structure and molecular mechanism. In *Nature* (Vol. 349, Issue 6305). https://doi.org/10.1038/349117a0

Bouvet, M., Lugari, A., Posthuma, C. C., Zevenhoven, J. C., Bernard, S., Betzi, S., Imbert, I., Canard, B., Guillemot, J. C., Lécine, P., Pfefferle, S., Drosten, C., Snijder, E. J., Decroly, E., & Morelli, X. (2014). Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. *Journal of Biological Chemistry*, *289*(37). https://doi.org/10.1074/jbc.M114.577353

Bryan, C. M., Bhandari, J., Napuli, A. J., Leibly, D. J., Choi, R., Kelley, A., van Voorhis, W. C., Edwards, T. E., & Stewart, L. J. (2011). High-throughput protein production and purification at the Seattle Structural Genomics Center for Infectious Disease. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, *67*(9). https://doi.org/10.1107/S1744309111018367

Burnett, M. J. B., & Burnett, A. C. (2020). Therapeutic recombinant protein production in plants: Challenges and opportunities. In *Plants People Planet* (Vol. 2, Issue 2). https://doi.org/10.1002/ppp3.10073

Butler, M., & Spearman, M. (2014). The choice of mammalian cell host and possibilities for glycosylation engineering. In *Current Opinion in Biotechnology* (Vol. 30). https://doi.org/10.1016/j.copbio.2014.06.010

Carlos Amor, J., Harrison, D. H., Kahn, R. A., & Ringe, D. (1994). Structure of the human ADP-ribosylation factor 1 complexed with GDP. *Nature*, *372*(6507). https://doi.org/10.1038/372704a0

Carter, P. J. (2011). Introduction to current and future protein therapeutics: A protein engineering perspective. In *Experimental Cell Research* (Vol. 317, Issue 9). https://doi.org/10.1016/j.yexcr.2011.02.013

Chen, T., Zhang, K., Gruenhagen, J., & Medley, C. D. (2015). Hydrophobic interaction chromatography for antibody drug conjugate drug distribution analysis. *American Pharmaceutical Review*, *18*(5).

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, *66*(1). https://doi.org/10.1107/S0907444909042073

Choi, J. H., & Lee, S. Y. (2004). Secretory and extracellular production of recombinant proteins using Escherichia coli. In *Applied Microbiology and Biotechnology* (Vol. 64, Issue 5). https://doi.org/10.1007/s00253-004-1559-9

Choi, R., Kelley, A., Leibly, D., Nakazawa Hewitt, S., Napuli, A., & van Voorhis, W. (2011). Immobilized metal-affinity chromatography protein-recovery screening is predictive of crystallographic structure success. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, *67*(9). https://doi.org/10.1107/S1744309111017374

Cohen, S. N., Chang, A. C. Y., Boyer, H. W., & Helling, R. B. (1973). Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, *70*(11). https://doi.org/10.1073/pnas.70.11.3240

Commandeur, U., & Twyman, R. M. (2005). Biosafety Aspects of Molecular Farming in Plants. In *Molecular Farming: Plant-made Pharmaceuticals and Technical Proteins*. https://doi.org/10.1002/3527603638.ch16

Crea, R., Kraszewski, A., Hirose, T., & Itakura, K. (1978). Chemical synthesis of genes for human insulin. *Proceedings of the National Academy of Sciences of the United States of America*, *75*(12). https://doi.org/10.1073/pnas.75.12.5765

Dill, K. A. (1990). Dominant Forces in Protein Folding. *Biochemistry*, *29*(31). https://doi.org/10.1021/bi00483a001

Donaldson, J. G., & Jackson, C. L. (2000). Regulators and effectors of the ARF GTPases. In *Current Opinion in Cell Biology* (Vol. 12, Issue 4). https://doi.org/10.1016/S0955-0674(00)00119-8

Dondapati, S. K., Stech, M., Zemella, A., & Kubick, S. (2020). Cell-Free Protein Synthesis: A Promising Option for Future Drug Development. In *BioDrugs* (Vol. 34, Issue 3). https://doi.org/10.1007/s40259-020-00417-y

Dove, A. (2002). Uncorking the biomanufacturing bottleneck. In *Nature Biotechnology* (Vol. 20, Issue 8). https://doi.org/10.1038/nbt0802-777

D'Souza-Schorey, C., & Chavrier, P. (2006). ARF proteins: Roles in membrane traffic and beyond. In *Nature Reviews Molecular Cell Biology* (Vol. 7, Issue 5). https://doi.org/10.1038/nrm1910

Eckerle, L. D., Becker, M. M., Halpin, R. A., Li, K., Venter, E., Lu, X., Scherbakova, S., Graham, R. L., Baric, R. S., Stockwell, T. B., Spiro, D. J., & Denison, M. R. (2010). Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathogens*, *6*(5). https://doi.org/10.1371/journal.ppat.1000896

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, *66*(4). https://doi.org/10.1107/S0907444910007493

Feng, Z., Li, X., Fan, B., Zhu, C., & Chen, Z. (2022). Maximizing the Production of Recombinant Proteins in Plants: From Transcription to Protein Stability. In *International Journal of Molecular Sciences* (Vol. 23, Issue 21). https://doi.org/10.3390/ijms232113516

Gengenbach, B. B., Keil, L. L., Opdensteinen, P., Müschen, C. R., Melmer, G., Lentzen, H., Bührmann, J., & Buyel, J. F. (2019). Comparison of microbial and transient expression (tobacco plants and plant-cell packs) for the production and purification of the anticancer mistletoe lectin viscumin. *Biotechnology and Bioengineering*, *116*(9). https://doi.org/10.1002/bit.27076

Gleba, Y. Y., Tusé, D., & Giritch, A. (2014). Plant viral vectors for delivery by Agrobacterium. *Current Topics in Microbiology and Immunology*, *375*. https://doi.org/10.1007/82-2013-352

Gomord, V., Chamberlain, P., Jefferis, R., & Faye, L. (2005). Biopharmaceutical production in plants: Problems, solutions and opportunities. In *Trends in Biotechnology* (Vol. 23, Issue 11). https://doi.org/10.1016/j.tibtech.2005.09.003

Gronemeyer, P., Ditz, R., & Strube, J. (2014). Trends in upstream and downstream process development for antibody manufacturing. In *Bioengineering* (Vol. 1, Issue 4). https://doi.org/10.3390/bioengineering1040188

Han, M., & Yu, X. (2015). Enhanced expression of heterologous proteins in yeast cells via the modification of N-glycosylation sites. *Bioengineered*, *6*(2). https://doi.org/10.1080/21655979.2015.1011031

Haque, R., Huston, C. D., Hughes, M., Houpt, E. R., & Petri, W. A. Jr. (2003). Amebiasis. Current Concepts. *New England Journal of Medicine*, *348*(16).

Häussinger, D., & Sies, H. (2007). Osmosensing and osmosignaling. *Methods in Enzymology, 428*.

He, Y., Ning, T., Xie, T., Qiu, Q., Zhang, L., Sun, Y., Jiang, D., Fu, K., Yin, F., Zhang, W., Shen, L., Wang, H., Li, J., Lin, Q., Sun, Y., Li, H., Zhu, Y., & Yang, D. (2011). Large-scale production of functional human serum albumin from transgenic rice seeds. *Proceedings of the National*

Academy of Sciences of the United States of America, *108*(47). https://doi.org/10.1073/pnas.1109736108

Irani, Z. A., Kerkhoven, E. J., Shojaosadati, S. A., & Nielsen, J. (2016). Genome-scale metabolic model of Pichia pastoris with native and humanized glycosylation of recombinant proteins. *Biotechnology and Bioengineering*, *113*(5). https://doi.org/10.1002/bit.25863

Jaenisch, R. (1976). Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. *Proceedings of the National Academy of Sciences of the United States of America*, *73*(4). https://doi.org/10.1073/pnas.73.4.1260

Jones, M. (2013). Sally Smith Hughes. Genentech: The Beginnings of Biotech. Chicago, IL: University of Chicago Press, 2011. 232 pp. ISBN 9780226045511, $16.00 (paper). *Enterprise and Society*, *14*(4). https://doi.org/10.1093/es/kht081

Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography*, *66*(2). https://doi.org/10.1107/S0907444909047374

Khan, K. H. (2013). Gene expression in mammalian cells and its applications. *Advanced Pharmaceutical Bulletin*, *3*(2). https://doi.org/10.5681/apb.2013.042

Kim, H. J., & Kim, H. J. (2017). Yeast as an expression system for producing virus-like particles: what factors do we need to consider? In *Letters in Applied Microbiology* (Vol. 64, Issue 2). https://doi.org/10.1111/lam.12695

Lee, S. Y. (1996). Bacterial polyhydroxyalkanoates. In *Biotechnology and Bioengineering* (Vol. 49, Issue 1). https://doi.org/10.1002/(SICI)1097-0290(19960105)49:1<1::AID-BIT1>3.3.CO;2-1

Lin, S., Chen, H., Chen, Z., Yang, F., Ye, F., Zheng, Y., Yang, J., Lin, X., Sun, H., Wang, L., Wen, A., Dong, H., Xiao, Q., Deng, D., Cao, Y., & Lu, G. (2021). Crystal structure of SARS-CoV-2 nsp10 bound to nsp14-ExoN domain reveals an exoribonuclease with both structural and functional integrity. *Nucleic Acids Research*, *49*(9). https://doi.org/10.1093/nar/gkab320

Lindsay, B. J., Bonar, M. M., Costas-Cancelas, I. N., Hunt, K., Makarkov, A. I., Chierzi, S., Krawczyk, C. M., Landry, N., Ward, B. J., & Rouiller, I. (2018). Morphological characterization of a plant-made virus-like particle vaccine bearing influenza virus hemagglutinins by electron microscopy. *Vaccine*, *36*(16). https://doi.org/10.1016/j.vaccine.2018.02.106

Looser, V., Bruhlmann, B., Bumbak, F., Stenger, C., Costa, M., Camattari, A., Fotiadis, D., & Kovar, K. (2014). Cultivation strategies to enhance productivity of Pichia pastoris: A review. In *Biotechnology Advances* (Vol. 33, Issue 6). https://doi.org/10.1016/j.biotechadv.2015.05.008

López-Mirabal, H. R., & Winther, J. R. (2008). Redox characteristics of the eukaryotic cytosol. In *Biochimica et Biophysica Acta - Molecular Cell Research* (Vol. 1783, Issue 4). https://doi.org/10.1016/j.bbamcr.2007.10.013

Malm, M., Diessner, A., Tamminen, K., Liebscher, M., Vesikari, T., & Blazevic, V. (2019). Rotavirus VP6 as an adjuvant for bivalent norovirus vaccine produced in nicotiana benthamiana. *Pharmaceutics*, *11*(5). https://doi.org/10.3390/pharmaceutics11050229

McCoy, A. J. (2006). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallographica Section D: Biological Crystallography*, *63*(1). https://doi.org/10.1107/S0907444906045975

Medrano, G., Reidy, M. J., Liu, J., Ayala, J., Dolan, M. C., & Cramer, C. L. (2009). Recombinant Proteins From Plants. *Proteins*, *483*. https://doi.org/10.1007/978-1-59745-407-0

Menkhaus, T. J., Bai, Y., Zhang, C., Nikolov, Z. L., & Glatz, C. E. (2004). Considerations for the recovery of recombinant proteins from plants. In *Biotechnology Progress* (Vol. 20, Issue 4). https://doi.org/10.1021/bp040011m

Moreira, M., & Sarraguça, M. (2020). How can oral paediatric formulations be improved? A challenge for the XXI century. In *International Journal of Pharmaceutics* (Vol. 590). https://doi.org/10.1016/j.ijpharm.2020.119905

Munoz, A. J., Wanichthanarak, K., Meza, E., & Petranovic, D. (2012). Systems biology of yeast cell death. In *FEMS Yeast Research* (Vol. 12, Issue 2). https://doi.org/10.1111/j.1567-1364.2011.00781.x

Myler, P., Stacy, R., Stewart, L., Staker, B., van Voorhis, W., Varani, G., & Buchko, G. (2009). The Seattle Structural Genomics Center for Infectious Disease (SSGCID). *Infectious Disorders - Drug Targets*, *9*(5). https://doi.org/10.2174/187152609789105687

Nasheuer, H. P., Smith, R., Bauerschmidt, C., Grosse, F., & Weisshart, K. (2002). Initiation of eukaryotic DNA replication: Regulation and mechanisms. *Progress in Nucleic Acid Research and Molecular Biology*, *72*. https://doi.org/10.1016/s0079-6603(02)72067-9

Nielsen, J. (2013). Production of biopharmaceutical proteins by yeast. *Bioengineered*, *4*(4). https://doi.org/10.4161/bioe.22856

Nosaki, S., Hoshikawa, K., Ezura, H., & Miura, K. (2021). Transient protein expression systems in plants and their applications. In *Plant Biotechnology* (Vol. 38, Issue 3). https://doi.org/10.5511/plantbiotechnology.21.0610a

Nowakowski, A. B., Wobig, W. J., & Petering, D. H. (2014). Native SDS-PAGE: High resolution electrophoretic separation of proteins with retention of native properties including bound metal ions. *Metallomics*, *6*(5). https://doi.org/10.1039/c4mt00033a

Paduch, M., Jeleń, F., & Otlewski, J. (2001). Structure of small G proteins and their regulators. In *Acta Biochimica Polonica* (Vol. 48, Issue 4). https://doi.org/10.18388/abp.2001_3850

Pasqualato, S., Renault, L., & Cherfils, J. (2002). Arf, Arl, Arp and Sar proteins: A family of GTP-binding proteins with a structural device for "front-back" communication. *EMBO Reports*, *3*(11). https://doi.org/10.1093/embo-reports/kvf221

Phan, H. T., Pham, V. T., Ho, T. T., Pham, N. B., Chu, H. H., Vu, T. H., Abdelwhab, E. M., Scheibner, D., Mettenleiter, T. C., Hanh, T. X., Meister, A., Gresch, U., & Conrad, U. (2020). Immunization with plant-derived multimeric H5 hemagglutinins protect chicken against highly pathogenic avian influenza virus H5N1. *Vaccines*, *8*(4). https://doi.org/10.3390/vaccines8040593

Phan, I. Q. H., Stacy, R., & Myler, P. J. (2014). Selecting targets from eukaryotic parasites for structural genomics and drug discovery. *Methods in Molecular Biology*, *1140*. https://doi.org/10.1007/978-1-4939-0354-2_4

Ratelade, J., Miot, M. C., Johnson, E., Betton, J. M., Mazodier, P., & Benaroudj, N. (2009). Production of recombinant proteins in the lon-deficient BL21(DE3) strain of Escherichia coli in the absence of the DnaK chaperone. *Applied and Environmental Microbiology*, *75*(11). https://doi.org/10.1128/AEM.00255-09

Rawlings, N. D., & Salvesen, G. (2013). Handbook of Proteolytic Enzymes. In *Handbook of Proteolytic Enzymes* (Vols. 1–3). https://doi.org/10.1016/C2009-1-60990-4

Razinkov, V. I., Treuheit, M. J., & Becker, G. W. (2015). Accelerated formulation development of monoclonal antibodies (MABS) and mab-based modalities: Review of methods and tools. In *Journal of Biomolecular Screening* (Vol. 20, Issue 4). https://doi.org/10.1177/1087057114565593

Ren, X., Farías, G. G., Canagarajah, B. J., Bonifacino, J. S., & Hurley, J. H. (2013). Structural basis for recruitment and activation of the AP-1 clathrin adaptor complex by Arf1. *Cell*, *152*(4). https://doi.org/10.1016/j.cell.2012.12.042

Ros, V. I. D. (2020). Baculoviruses: General Features (Baculoviridae). In *Encyclopedia of Virology: Volume 1-5, Fourth Edition* (Vols. 1–5). https://doi.org/10.1016/B978-0-12-809633-8.21549-5

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. In *Frontiers in Microbiology* (Vol. 5, Issue APR). https://doi.org/10.3389/fmicb.2014.00172

Royal, J. M., Simpson, C. A., McCormick, A. A., Phillips, A., Hume, S., Morton, J., Shepherd, J., Oh, Y., Swope, K., Debeauchamp, J. L., Webby, R. J., Cross, R. W., Borisevich, V., Geisbert, T. W., Demarco, J. K., Bratcher, B., Haydon, H., & Pogue, G. P. (2021). Development of a sars-cov-2 vaccine candidate using plant-based manufacturing and a tobacco mosaic virus-like nano-particle. *Vaccines*, *9*(11). https://doi.org/10.3390/vaccines9111347

Sahdev, S., Khattar, S. K., & Saini, K. S. (2008). Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies. In

Molecular and Cellular Biochemistry (Vol. 307, Issues 1–2). https://doi.org/10.1007/s11010-007-9603-6

Schmidt, T. G. M., & Skerra, A. (2007). The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nature Protocols*, *2*(6). https://doi.org/10.1038/nprot.2007.209

Serbzhinskiy, D. A., Clifton, M. C., Sankaran, B., Staker, B. L., Edwards, T. E., & Myler, P. J. (2015). Structure of an ADP-ribosylation factor, ARF1, from Entamoeba histolytica bound to Mg2+-GDP. *Acta Crystallographica Section F: Structural Biology Communications*, *71*. https://doi.org/10.1107/S2053230X15004677

Sezonov, G., Joseleau-Petit, D., & D'Ari, R. (2007). Escherichia coli physiology in Luria-Bertani broth. *Journal of Bacteriology*, *189*(23). https://doi.org/10.1128/JB.01368-07

Shanmugaraj, B., Bulaon, C. J. I., & Phoolcharoen, W. (2020). Plant molecular farming: A viable platform for recombinant biopharmaceutical production. In *Plants* (Vol. 9, Issue 7). https://doi.org/10.3390/plants9070842

Shiloach, J., & Fass, R. (2005). Growing E. coli to high cell density - A historical perspective on method development. In *Biotechnology Advances* (Vol. 23, Issue 5). https://doi.org/10.1016/j.biotechadv.2005.04.004

Shoji, Y., Farrance, C. E., Bautista, J., Bi, H., Musiychuk, K., Horsey, A., Park, H., Jaje, J., Green, B. J., Shamloul, M., Sharma, S., Chichester, J. A., Mett, V., & Yusibov, V. (2012). A plant-based system for rapid production of influenza vaccine antigens. *Influenza and Other Respiratory Viruses*, *6*(3). https://doi.org/10.1111/j.1750-2659.2011.00295.x

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*. https://doi.org/10.1038/msb.2011.75

Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M., & Panda, A. K. (2015). Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. In *Microbial Cell Factories* (Vol. 14, Issue 1). https://doi.org/10.1186/s12934-015-0222-8

Sprang, S. R. (1997). G protein mechanisms: Insights from structural analysis. In *Annual Review of Biochemistry* (Vol. 66). https://doi.org/10.1146/annurev.biochem.66.1.639

Streatfield, S. J. (2007). Approaches to achieve high-level heterologous protein production in plants. In *Plant Biotechnology Journal* (Vol. 5, Issue 1). https://doi.org/10.1111/j.1467-7652.2006.00216.x

Tanwar, J., Das, S., Fatima, Z., & Hameed, S. (2014). Multidrug resistance: An emerging crisis. In *Interdisciplinary Perspectives on Infectious Diseases* (Vol. 2014). https://doi.org/10.1155/2014/541340

Tekoah, Y., Shulman, A., Kizhner, T., Ruderfer, I., Fux, L., Nataf, Y., Bartfeld, D., Ariel, T., Gingis-Velitski, S., Hanania, U., & Shaaltiel, Y. (2015). Large-scale production of pharmaceutical proteins in plant cell culture-the protalix experience. In *Plant Biotechnology Journal* (Vol. 13, Issue 8). https://doi.org/10.1111/pbi.12428

Tregoning, J. S. (2020). First human efficacy study of a plant-derived influenza vaccine. In *The Lancet* (Vol. 396, Issue 10261). https://doi.org/10.1016/S0140-6736(20)32010-9

Vetter, I. R., & Wittinghofer, A. (2001). The guanine nucleotide-binding switch in three dimensions. In *Science* (Vol. 294, Issue 5545). https://doi.org/10.1126/science.1062023

Ward, B. J., Séguin, A., Couillard, J., Trépanier, S., & Landry, N. (2021). Phase III: Randomized observer-blind trial to evaluate lot-to-lot consistency of a new plant-derived quadrivalent virus like particle influenza vaccine in adults 18–49 years of age. *Vaccine*, *39*(10). https://doi.org/10.1016/j.vaccine.2021.01.004

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. In *Acta Crystallographica Section D: Biological Crystallography* (Vol. 67, Issue 4). https://doi.org/10.1107/S0907444910045749

Wuest, D. M., Harcum, S. W., & Lee, K. H. (2012). Genomics in mammalian cell culture bioprocessing. In *Biotechnology Advances* (Vol. 30, Issue 3). https://doi.org/10.1016/j.biotechadv.2011.10.010

Wurm, F. M. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. In *Nature Biotechnology* (Vol. 22, Issue 11). https://doi.org/10.1038/nbt1026

Xu, S., Gavin, J., Jiang, R., & Chen, H. (2017). Bioreactor productivity and media cost comparison for different intensified cell culture processes. *Biotechnology Progress*, *33*(4). https://doi.org/10.1002/btpr.2415

Zha, J., Liu, D., Ren, J., Liu, Z., & Wu, X. (2023). Advances in Metabolic Engineering of Pichia pastoris Strains as Powerful Cell Factories. In *Journal of Fungi* (Vol. 9, Issue 10). https://doi.org/10.3390/jof9101027

Zimmerman, R. K. (2007). HPV vaccine and its recommendations, 2007. In *Journal of Family Practice* (Vol. 56, Issue SUPPL. 2).

Zimran, A., Brill-Almon, E., Chertkoff, R., Petakov, M., Blanco-Favela, F., Muñoz, E. T., Solorio-Meza, S. E., Amato, D., Duran, G., Giona, F., Heitner, R., Rosenbaum, H., Giraldo, P., Mehta, A., Park, G., Phillips, M., Elstein, D., Altarescu, G., Szleifer, M., … Aviezer, D. (2011). Pivotal trial with plant cell-expressed recombinant glucocerebrosidase, taliglucerase alfa, a novel enzyme replacement therapy for Gaucher disease. *Blood*, *118*(22). https://doi.org/10.1182/blood-2011-07-366955